
DIAGNOSING
PERSONALITY
AND
CONDUCT

The Century Psychology Series
Richard M. Elliott, Editor

DIAGNOSING PERSONALITY AND CONDUCT

By
PERCIVAL M. SYMONDS

ASSOCIATE PROFESSOR OF EDUCATION
TEACHERS COLLEGE, COLUMBIA UNIVERSITY
AUTHOR OF "THE NATURE OF CONDUCT"



APPLETON-CENTURY-CROFTS, INC.
NEW YORK

COPYRIGHT, 1931, BY THE CENTURY CO.
ALL RIGHTS RESERVED, INCLUDING THE
RIGHT TO REPRODUCE THIS BOOK, OR POR-
TIONS THEREOF, IN ANY FORM. 560

PREFACE

Progress in the control of inanimate nature has been accomplished in part by the development of more exact methods of describing natural phenomena. This control of nature has been particularly facilitated by the invention of instruments of measurement. Similarly the control of human conduct and education depends on the development of more exact methods of describing human conduct. The exact description of human conduct can be rendered most efficient when it is reduced to a form of measurement, for then small differences are most accurately portrayed and small changes most accurately noted.

Treatises on psychological measurement have in the past dealt almost entirely with human *abilities*. Tests and scales have been devised for measuring the general level of intellectual ability, achievement in school subjects, and aptitude in various special abilities. As these measures have been applied to the measurement of products of learning in school and to the estimation of capacity in the workaday world, it has become evident that certain important educational outcomes were being neglected. Tests might show how well a pupil is achieving in subject-matter, but would not at all show what interests were developing, what attitudes were being formed, the adequacy of his personal adjustments, or the development of character. A whole side of human nature has been neglected in the development of tests and scales of ability.

Elated with their success with the measurement of intelligence during the World War, psychologists turned their attention to the measurement of personality and character. In a symposium on *Intelligence and Its Measurement* in 1921 many of the contributors stated that one of the next steps in research was the development of measurement of character. As the late Dr. Colvin put it, "The most important 'next step' for purposes both of prognosis and diagnosis is the formulation of a test that will inform us of the character qualities of those tested." Pintner stated: "I feel

that the time is now ripe for active investigation of the emotions, the character, the will, and so forth, by means of mental test methods." Pressey said: "There simply must be a courageous attack upon the problem of measurement of other than intellectual factors. It is becoming increasingly obvious that matters of temperament and character are of very great importance, that they operate quite largely independent of intelligence, that prognosis problems cannot be adequately understood without an evaluation of these factors." Terman stated as one of the next steps in research "investigations of instinctive, emotional and volitional traits and of the combinations of these which are involved in prepsychopathic conditions and normal variations in temperament." Thurstone said, "I should like to see another line of mental test work opened up, namely the diagnosis of the volitional and emotional characteristics which determine our character traits."

Many of the methods and techniques described in this book have been patiently and quietly investigated in the psychological laboratory for many years. But within the past decade all have been applied to practical problems of human conduct with the result that earlier methods have been refined and extended. They are now being tried experimentally or are in practical use in the work of the school, industry, clinic, court, and wherever else there are problems of human adjustment. This book summarizes the research that has been done in these various methods and techniques. Since the writer approaches the problem with a background in educational psychology, the viewpoint, illustrations, and applications tend to be largely educational. The book is primarily addressed to educators whose breadth of view causes them to see conduct as a primary concern of education. But since practical interest in the techniques described in this book is by no means limited to educators, any social scientist, whether in medicine, industry, law, or religion, may find herein discussed methods and techniques for the more exact study of human relationships.

Several threads of thought run through the discussion of the various methods which tend to give the whole a unified point of view. The refinement of any technique indicates the need for adequate sampling. The writer believes that progress in the diagnosis of conduct inevitably leads toward standardization, and he has been at pains to point out the possibility and advantages of standardization wherever possible. The empirical approach to

the study of conduct must perforce be kept uppermost. Though guided by the best cues and hypotheses available, any investigator must plead ignorance as he delves into the unknown, and must depend on results to check theory and dogmatism. At every turn pains will be taken to stress the fallibility of naked human judgment in diagnosis. If there be any one guiding principle that rises above the others, it is the resolve to make use of relationships adequately proved by careful experimentation instead of depending on limited and unsystematic personal experience.

The writer remembers hearing the measurement of non-intellectual qualities discussed years ago as though there were unlimited possibilities in the development of "character tests." But he is bold enough now to assert that the main lines which such "testing" must follow are included in the chapter headings of this book. Certain fields, to be sure, have been omitted. Hypnosis has potentiality both as a diagnostic and as a therapeutic technique, but its development in this direction has been very rudimentary. The study of personal letters, diaries, essays, and autobiographies is still another source of evidence, the value of which has yet to be determined. On the other hand the methods herein described involve measurement of the environment as it conditions conduct, measurement of the actual conduct itself, and measurement of the results of conduct; they involve direct observation and interpretation of what is observed, and the testimony of others than the observer concerning the findings; and they involve the direct study of reactions and verbal testimony concerning reactions. That there must be some sort of evidence in the accurate description of conduct is indisputable, and this evidence must proceed from one or the other of the above sources. But there is infinite possibility for development and refinement.

An explanation should be made as to what this book does not include as well as to what it does include. Tests of ability and intelligence have been carefully avoided, not because they are without significance in the diagnosis of conduct, but because they are adequately described and discussed elsewhere. In several places, particularly in the chapters on interviewing, psychoanalysis, and case studies, it was difficult to avoid a discussion of the treatment and therapeutic value of these techniques. However, since diagnosis and treatment can be distinguished, so far as possible discussion has been confined to diagnosis. It would be impossible to do

justice to the therapeutic values of these techniques within the limits of this volume.

The manuscript of this book was started in 1927. Since that time the remarkable *Studies in the Nature of Character* of the Character Education Inquiry have appeared. These studies are fundamental to an understanding of the nature of conduct and to the future development of methods of investigation. The writer wishes to express his debt to the work of Hartshorne and May, which is only too inadequately represented in this book. Space does not permit acknowledgment at this point to the many investigators whose researches have thrown light on the problems discussed, but reference is given to their work throughout the text. Certain contributors stand out as deserving special mention here: D. S. Thomas and Willard Olson for the development of techniques of observation; Woodworth for the original psychoneurotic inventory; Goodwin Watson, for both his own and his students' contributions to the measurement and diagnosis of attitudes and adjustment; and finally Freud, whose development of psychoanalysis has influenced the techniques of diagnosing conduct in more ways than can be easily described.

The reference lists at the end of each chapter contain most of the important references on each topic through 1929. Any one planning to do further work on a topic should make a careful survey of work done in the interval after that date.

The writer is under considerable obligation to Dr. E. L. Thorndike for reading the manuscript and for helpful criticism. Dr. D. G. Paterson and Dr. Florence Goodenough have made valuable suggestions. Above all appreciation is due to the general editor of the Century Psychology Series, Dr. R. M. Elliott, for expert criticism and advice.

CONTENTS

CHAPTER	PAGE
I INTRODUCTION: SIGNIFICANCE OF THE DIAGNOSIS OF CONDUCT	3
II OBSERVATION	23
III RATING METHODS	41
IV THE QUESTIONNAIRE	122
V ADJUSTMENT QUESTIONNAIRES	174
VI ATTITUDE QUESTIONNAIRES	215
VII INTEREST QUESTIONNAIRES	239
VIII TESTS OF CONDUCT, KNOWLEDGE, AND JUDGMENT . .	260
IX PERFORMANCE TESTS	298
X THE FREE ASSOCIATION METHOD	361
XI PHYSIOLOGICAL MEASURES OF THE EMOTIONS . .	400
XII INTERVIEWING	450
XIII PSYCHOANALYSIS	485
XIV EXTERNAL SIGNS OF CONDUCT	503
XV MEASURES OF THE ENVIRONMENT	536
XVI THE CASE STUDY: A COMPREHENSIVE STUDY OF THE INDIVIDUAL	555
BIBLIOGRAPHY	571
INDEX	573

LIST OF ILLUSTRATIONS

	PAGE
SCALE OF ATTITUDES	231
CARDIOBOARD TEST	305
THE DUDGEON SPHYGMOGRAPH IN POSITION	404
SPHYGMOMANOMETER	408
PLETHYSMOGRAPH	416
PNEUMOGRAPH	418
D'ARSONVAL GALVANOMETER	422
ELECTRIC CONNECTIONS FOR APPARATUS FOR MEASURING THE PSYCHOGALVANIC REFLEX	423
CONDUCT IN SPECIFIC SITUATIONS	565

LIST OF TABLES

TABLE	PAGE
1 CONSTANCY OF ORAL NERVOUS HABITS OVER AN INTERVAL	35
2 RELIABILITY COEFFICIENTS FOR OBSERVATION	36
3 COMPUTATION FOR DETERMINING AVERAGE ERROR AND DIRECTION OF ERROR IN RATING	48
4 RELIABILITY COEFFICIENTS FROM PERSONAL ATTITUDES TEST SELF-ORDINARY—IDEAL RATING	76
5 DISTRIBUTION SHOWING NUMBER OF SCALE DIVISIONS IN 54 RATING SCALES	78
6 NUMBER OF CLASSES FOR RATING SCALES YIELDING DIFFERENT RELIABILITIES	80
7 PERCENTAGE DISTRIBUTIONS FOR GROUPS OF DIFFERENT SIZES AND FOR SCALES HAVING DIFFERENT NUMBERS OF CLASSES	81
8 THE AMOUNTS OF DIFFERENCE (IN y) CORRESPONDING TO GIVEN PERCENTAGE OF JUDGMENTS THAT $x > y$	87
9 PER CENT OF JUDGES WHO BELIEVE ONE ITEM TO BE LESS VALUABLE THAN ANOTHER	88
10 PERCENTAGES OF TABLE 9 TRANSMUTED INTO CORRESPONDING STANDARD DEVIATION VALUES	88
11 TABLE OF STANDARD DEVIATIONS TO BE ASSIGNED FOR RANK POSITIONS	90
12 TABLE FOR TRANSMUTING RANKINGS INTO UNITS OF AMOUNT	91
13 THE RANKINGS OF EIGHT PUPILS TURNED INTO UNITS OF AMOUNTS	92
14 NUMBER OF RATINGS NECESSARY TO OBTAIN A SPECIFIED RELIABILITY FOR VARIOUS TRAITS	96
15 RELIABILITY COEFFICIENTS OF RATINGS FOR TEN TRAITS	104
16 AMOUNT OF DISAGREEMENT AMONG JUDGES IN ESTIMATING, ON THE BASIS OF ACQUAINTANCE, THE TRAITS OF OTHERS, IN TWO INVERTATIONS	105
17 AVERAGE DEVIATION OF TRAITS IN RANKING	105
18 RELIABILITY COEFFICIENTS OF RATINGS FOR SIX TRAITS	106
19 RELIABILITY COEFFICIENTS OF RATING EIGHT TRAITS	106
20 PERCENTAGE OF TIMES A SUBJECT FOLLOWED THE LEAD OF A QUESTION	137
21 PERCENTAGE OF ACCURACY IN ANSWERING A QUESTIONNAIRE	164
22 RELIABILITY COEFFICIENTS OF QUESTIONNAIRES	166
23 RELIABILITY OF TYPICAL OBJECTIVE SUBJECT-MATTER TESTS	168
24 CLASSIFICATION OF ITEMS IN THE WOODWORTH PSYCHONEUROTIC INVENTORY	178
25 SECTIONS OF SYMONDS ADJUSTMENT QUESTIONNAIRE	182
26 RELIABILITY COEFFICIENTS—CHAMBERLAIN EXPERIENCE VARIABLES RECORD	185
27 FREEMAN X-O TESTS—RELIABILITY COEFFICIENTS	192
28 CORRELATIONS BETWEEN FREEMAN X-O TEST AND INTELLIGENCE AND SCHOOL MARKS	195

TABLE	PAGE
29 FREQUENCY DISTRIBUTION OF SCORES ON INTROVERSION-EXTROVERSION QUESTIONNAIRE	200
30 GROUP DIFFERENCES IN EXTROVERSION-INTROVERSION	203
31 BIALLELIAL r BETWEEN INTROVERSION-EXTROVERSION DIFFERENCE AND VARIOUS FACTORS	205
32 EVALUATION OF SINGLE ITEM ON ALLPORT ASCENDANCE-SUBMISSION QUESTIONNAIRE SHOWING DEVIATION OF SCORING SCHEME	206
33 INTERCORRELATIONS BETWEEN THURSTONE NEOTEST INVENTORY, BEAN-REUTER SELF-SUFFICIENCY TEST, LAIRD G. 2 INTROVERSION TEST, ALLPORT ASCENDANCE-SUBMISSION REACTION STUDY	209
34 RATINGS OF ARGUMENTS	220
35 INTERCORRELATIONS OF SCHOOL MARKS, RATINGS FOR STUDIOUSNESS, STUDIOUSNESS INDEX, AND STUDIOUSNESS QUESTIONNAIRE	227
36 RELIABILITY OF ATTITUDES QUESTIONNAIRES	232
37 CORRELATIONS BETWEEN GROUP SCORES FOR EACH TEST FORM AND GROUP SCORES FOR TOTAL TEST	234
38 PROBABLE PERMANENCY OF VOCATIONAL AMBITION BETWEEN VARIOUS DEVELOPMENTAL PERIODS	241
39 PROBABLE RELATION OF INTELLIGENCE AND INTELLIGENCE REQUIREMENT OF VOCATIONAL AMBITION AT VARIOUS DEVELOPMENTAL PERIODS	243
40 CORRELATION OF INTEREST IN SCHOOL SUBJECTS AND ACHIEVEMENT	244
41 GARRETTSON'S INTEREST QUESTIONNAIRE (SECTIONS AND NUMBERS OF ITEMS)	250
42 DISTRIBUTIONS FOR THREE SCHOOLS ON GARRETTSON QUESTIONNAIRE SCORED FOR COMMERCIAL INTEREST	251
43 RELIABILITY OF FRAUD INTEREST BLANK	252
44 SHOWING CHANGES IN SCORE FROM ORIGINAL TEST (FRAUD'S INTEREST ANALYSIS BLANK) TO RETEST	252
45 RELIABILITY COEFFICIENTS OF VOCATIONAL INTEREST BLANK	253
46 PERCENTAGES OF MEN IN VARIOUS OCCUPATIONS WHO RATE <i>A</i> AND <i>B</i> IN THE INTEREST OF CERTIFIED PUBLIC ACCOUNTANTS, ENGINEERS, AND PERSONNEL MANAGERS	254
47 CORRELATIONS BETWEEN SCORES OBTAINED BY CERTIFIED PUBLIC ACCOUNTANTS, ENGINEERS, AND PERSONNEL MANAGERS IN THEIR OWN AND OTHER OCCUPATIONAL INTERESTS	254
48 RELIABILITY COEFFICIENTS OF GARRETTSON INTEREST QUESTIONNAIRE FOR HIGH SCHOOL STUDENTS	255
49 SCORING JUDGMENTS EXPRESSED AS RATINGS	262
50 RELIABILITY COEFFICIENTS OF TEST OF CONDUCT INFORMATION AND OPINION	264
51 RELIABILITY COEFFICIENTS—MORAL KNOWLEDGE TESTS	265
52 INTERCORRELATIONS OF I. E. R. MORAL KNOWLEDGE TESTS	266
53 CORRELATION OF JUDGMENT AND REASONING TESTS AGAINST A COMPOSITE OF VOCABULARY AND KNOWLEDGE TESTS	267
54 CORRELATION OF MORAL KNOWLEDGE AND INTELLIGENCE	268
55 CORRELATION BETWEEN MORAL KNOWLEDGE TESTS AND AGE	268
56 RELATION OF CHEATING TO MORAL KNOWLEDGE	270
57 CORRELATIONS BETWEEN MORAL KNOWLEDGE OF CHILDREN AND OF OTHER GROUPS	271
58 CORRELATIONS BETWEEN MORAL KNOWLEDGE AND CONDUCT FOR INDIVIDUALS AND GROUPS—AVERAGE OF EIGHT TESTS	273

List of Tables

xv

TABLE	PAGE
59 INTERCORRELATIONS OF CIRCULAR SPACES AND MAZE TESTS	307
60 CORRELATIONS OF CIRCULAR SPACES AND MAZE TESTS WITH INCORPORABILITY	307
61 RELIABILITY OF HONESTY DIFFICULTY TESTS	312
62 RELIABILITY OF HONESTY STEEL TESTS	312
63 RELIABILITY OF TYPES OF DECEPTIVE BEHAVIOR	316
64 INTERCORRELATIONS OF NINE TYPES OF DECEPTIVE BEHAVIOR	317
65 AVERAGE INTERCORRELATIONS BETWEEN SINGLE TESTS OF DIFFERENT TECHNIQUES	317
66 RELIABILITY OF PERSISTENCE TESTS	325
67 INTERCORRELATIONS OF PERSISTENCE TESTS	325
68 INTERCORRELATIONS OF SERVICE TESTS	326
69 INTERCORRELATIONS OF INHIBITION TESTS	327
70 CORRELATIONS BETWEEN THE DIFFERENT MEASURES FOR SPEED OF DECISION	330
71 RELIABILITY COEFFICIENTS ON THE GROUP TEST, DOWNEY WILL-TEMPERAMENT TEST	341
72 INTERCORRELATIONS OF TESTS OF SPEED OF MOVEMENT	345
73 AVERAGE INTERCORRELATION OF THE DOWNEY WILL-TEMPERAMENT TESTS	346
74 CORRELATIONS BETWEEN SCORES ON THE DOWNEY WILL-TEMPERAMENT TEST AND RATINGS FOR THE SAME TRAITS	348
75 KENT-BROWNELL LIST OF FREE-ASSOCIATION WORDS	362
76 JENK LIST (MODIFIED BY EMM) OF FREE-ASSOCIATION WORDS	363
77 CORRELATIONS BETWEEN COMPLEX SENS IN FREE ASSOCIATION	380
78 PERCENTAGES OF NORMAL AND IRVINE SUBJECTS SHOWING COMMON, DOMINANT, AND INDIVIDUAL RESPONSES IN FREE ASSOCIATION	386
79 PULSE RATES AT VARIOUS AGES	405
80 BLOOD-PRESSURES AT VARIOUS AGES	412
81 RATE OF BREATHING AT VARIOUS AGES	419
82 INSPIRATION-EXPIRATION RATION FOR VARIOUS STIMULI	435
83 CORRELATIONS OF ALKALINITY OF SALIVA AND ACIDITY OF URINE WITH PERSONALITY RATINGS	437
84 RARE ASSOCIATED AFFECTIONS BY INSTRUMENTS	478
85 RELATION BETWEEN TEMPERAMENTAL TYPES AND PHYSIQUE	511
86 SHOWING THE PERCENTAGE DISTRIBUTION OF MANIC-DEPRESSIVES AND SCHIZOPHRENICS ACCORDING TO KRETZSCHMAR'S BODY TYPES	512
87 CORRELATION BETWEEN RATINGS OF CHARACTER FACTORS AND PHYSIOLOGIC MEASUREMENTS ALLEGED TO BE SYMPTOMATIC OF THESE FACTORS	518
88 CORRELATIONS BETWEEN MEASUREMENTS OF FACIAL PROFILE AND PERSONALITY RATINGS	520
89 CORRELATIONS BETWEEN REAL CHARACTER TRAITS AND JUDGMENTS OF CHARACTER BASED ON PHOTOGRAPHS	521
90 PERCENTAGES OF DIPLOMS AND BACCALAUREATES RATED AS POSSESSING CERTAIN PERSONALITY TRAITS	523
91 CORRELATIONS BETWEEN CERTAIN MEASURED PROPORTIONS OF THE HAND AND THE CHARACTER TRAITS ALLEGED TO BE INDICATED BY EACH	524

List of Tables

TABLE	PAGE
92 CLAIMS MADE OF RELATIONSHIP BETWEEN CHARACTER TRAITS AND HANDWRITING CHARACTERISTICS	517
93 CORRELATIONS OF RATINGS OF CHARACTER AND HANDWRITING	518
94 CORRELATION BETWEEN RATINGS OF CHARACTER TRAITS AND HANDWRITING	518
95 OCCUPATIONAL INTELLIGENCE STANDARDS BASED ON ARMY ALPHA INTELLIGENCE TESTS	519
96 BI-SERIAL CORRELATION COEFFICIENTS BETWEEN ASPECTS OF HOME BACKGROUND AND SOCIO-ECONOMIC STATUS OF THE HOME	552
97 RELATIVE SIGNIFICANCE OF PERFORMANCE FACTORS FOR TOTAL REPUTATION	561
98 RELATIVE SIGNIFICANCE FOR CHARACTER OF MISCELLANEOUS CONCOMITANTS	563
99 CORRELATION OF GENERAL INTEGRATION WITH MISCELLANEOUS MEASURES	564

**DIAGNOSING PERSONALITY AND
CONDUCT**

DIAGNOSING PERSONALITY AND CONDUCT

Chapter I

INTRODUCTION: SIGNIFICANCE OF THE DIAGNOSIS OF CONDUCT

THE term *diagnosis*, as used in this book, has a variety of meanings. At one extreme it means mere measurement or description—the description of a person's behavior, or of a person's interests, or of his home. When used in this simple sense, *diagnosis* is a synonym for *survey*. At other times *diagnosis* means *discovery*—the discovery of individuals possessing certain characteristics. We speak of the discovery of predelinquents, or of problem children, and insofar as this discovery involves the accurate measurement or description of certain characteristics of these individuals, it too becomes a form of diagnosis.

In a narrower and more specific sense, *diagnosis* refers to the more complete understanding of an individual and of his conduct. If a problem child in school has already been singled out for study, then to proceed with diagnosis is to ascertain further facts concerning him, his home environment, his companions, his wishes and desires, etc., always with a view to fathoming the particular circumstance or combination of circumstances which seems to be responsible for his condition. In short, the teacher or psychologist wants to know *why* he is a problem child and what can be done about it.

To solve this problem two sets of facts are needed. One must know *first* the facts with regard to the child and *second* the significance of such facts as have a bearing on the situation in question, as distinguished from those which are merely the chance concomitants. For instance, if a child is caught stealing, we want to know many general things about both him and his environment. Then we also want to know the relationship of these factors in the child's life to stealing so that their significance in the particular situation may be determined.

4 Diagnosing Personality and Conduct

Nature of Measurement of Conduct

In a discussion of human conduct many persons are confused by the reference to measurement as a preferred form of diagnosis. The most obvious evidence that can be gathered with respect to human conduct is purely descriptive. Of a pupil in school it may be said, for example, that he has a clean face and clean hands, that he is courteous in class and on the playground, that he enjoys participating in group activities, that he has no brothers and sisters, or that he is trustworthy and industrious. Much of the clinical diagnosis resulting in the case study carried on in schools and in social work consists of this accurate description of the occurrence of isolated events for which there is no need for measurement.

But mere description of this simple sort does not satisfy all of the functions of diagnosis. One is not interested merely in knowing that John came to school on Tuesday morning and tried to bluff through a recitation in his history class for which he was not prepared. One wants to know how frequently this sort of behavior occurs in his various classes, whether there is a similar tendency to bluff outside of school, and whether he exhibits other symptoms indicating that he is not in proper adjustment to the work of the school.

Again, merely to know that a man expressed his opinion that England does not sincerely want naval reduction, but wishes to gain an advantage over other countries, especially the United States, in the disarmament conference, informs us very little about the man. We should want to know his reactions to many different kinds of social issues before forming any sort of impression as to his fair-mindedness and his tendencies toward liberalism or conservatism. From a single isolated event one is powerless to predict the future. In order to pick a satisfactory employee, a good wife, a pupil showing delinquent tendencies, a pupil showing personality disturbances, or a good city mayor, one needs to know certain trends in the individual, at least enough to be able to predict with some assurance what to expect from the individual in a new situation.

Thorndike has wisely said, "Whatever exists exists in some amount and can be measured." As we pass judgments on people in daily life we are constantly making these measurements in a

rough way. As we call this man honest, that man prompt, another well-balanced, another vain, and another introverted, we are making very crude measurements in the form of judgments. Hartshorne and May found that it was exceedingly rough and unsatisfactory to label people merely as honest or dishonest. One wants to know whether a man is honest in this, that, or the other situation, and under various degrees of temptation on several different occasions. Measurement merely attempts to make these descriptions finer and more exact. Instead of the two classifications *honest* and *dishonest*, the scale might run from *never honest*, through *seldom*, *sometimes*, and *usually*, to *always honest*. Or the findings might be "honest in fifteen out of twenty situations in which there was opportunity to steal, cheat, or lie." In proportion as the gradations become finer, one leaves qualitative description and adopts a form of measurement. The perfecting of measurement requires that attention be paid to the equality and evenness of units and to the zero point of the scale.

A key-note of this book, a point of emphasis to crop out again and again in the discussion of various techniques of investigation, is the emphasis on *adequacy of sampling*. A single observation is unreliable, a single rating is unreliable, a single test is unreliable, a single measurement is unreliable, a single answer to a question is unreliable. Reliability is achieved by heaping up observations, ratings, tests, questions, measures. Ask a boy to-day what he wants to do when he grows up, and he may respond, "Be an aviator," when a few years ago his choice would have been "Locomotive engineer." Because a boy's choices tend to be fickle and to change even week by week, we have him indicate his liking or disliking for each of a hundred or so occupations, activities, studies, and people. Then the outlines of certain broad, fundamental interests emerge which are stable and important. If you ask one teacher for her judgment of a boy's trustworthiness, you obtain what she has been able to observe in those few narrow class-room situations that appeared when her attention was particularly directed to some act involving honesty. An adequate rating, on the other hand, requires the judgment of several raters in several situations at several different times. Reliable evidence must be multiplied evidence.

Two main streams of investigation have developed in the field of conduct diagnosis. One may be called the *experimental ap-*

6 Diagnosing Personality and Conduct

proach. Investigators using experimental techniques have in great part been interested in questions as to the nature of behavior and character, and of the forces affecting the development of personality. In general these investigators have been trained in laboratory techniques or in psychological testing methods and have employed statistical methods in the validation of the diagnostic techniques. The experimentalist's insistence on isolating the effect of a single variable, on eliminating or at least controlling extraneous factors, and on the value of large samplings has played an important part in the development of these diagnostic techniques.

The other line of approach to the problem of diagnosing conduct may be called the *clinical*. Since the primary purpose of investigators using clinical methods has been to render *service*, they have been more concerned with using the techniques in guidance, in curing the sick, in placing people in industry, and in keeping them out of conflict with society, than in studying techniques analytically and ascertaining relationships between phenomena. These investigators have developed techniques that could be immediately used in a practical way in the work of the clinic, employment office, school, or court. Rating methods, the standardized interview, and psychoanalysis are typical products of this group of workers, and while these do not equal the questionnaire or testing methods in respect to accuracy, they have a demonstrable practical value.

These two groups of investigators have not always understood each other, and therefore have often worked at cross purposes. Clinical workers, desiring devices that could be used immediately, have failed to test these devices thoroughly before using them and have also lacked caution when interpreting results. Because their emphasis is concentrated upon interpreting a particular present situation, they tend to substitute for the painstaking analysis of statistical inquiry judgments and inferences based on clinical experience. Indeed, the clinical worker distrusts the conclusions drawn from statistics; they seem to have a sort of unreality for him, not inviting the confidence he places on things observed with his own eyes, felt with his own hands, or heard with his own ears. Then, too, he harbors the belief that statistics cover up differences, that since every individual presents a unique problem, is in fact a new bundle of forces existing in a combination that never existed

together before, statistical methods are too clumsy to deal with individuals. To him nothing can be better than the naked human judgment trained by adequate clinical experience.

In answer to this the experimentalist points out the fallacy of drawing conclusions from a single case or even from a small number of cases. The clinical worker may notice that a criminal has a peculiarly shaped skull and then conclude that all criminals have some sort of brain deformity. The clinical worker may notice that some pupils in school who have reading difficulties have more or less severe cases of "congenital word blindness" and build on this observation a theory of reading disability. Some gifted children having been observed to be weak and puny, there arises the popular generalization that bright children are frail and sickly. These observations become either trivial or false when large numbers of children are accurately observed under comparable conditions.

For such reasons the experimentalist is extremely skeptical of the value of the case study method for arriving at general truths based on observation and interviews. To him the experience gained from first-hand contact with actual cases has only a small chance of leading to valid generalizations unless, as is unlikely, this experience has been based on a large number of cases under carefully controlled conditions where accurate observation or measurement is possible.

A retort from the clinical worker may charge the experimentalist with merely talking theory and belittle his tests and measures because they do not begin to furnish all of the evidence needed in treating a practical case and are too clumsy or time-consuming even when they are available.

In conclusion it must be conceded that valuable diagnostic techniques have been developed by each school. To the clinical worker we are indebted for techniques which describe the present situation and for promising suggestions and challenging hypotheses regarding the significance of these facts. To the experimentalist we are indebted for the exhaustive, laborious checking-up of these hypotheses and the development of tested and standardized diagnostic devices for measuring them. To the experimentalist we must ultimately look for the accelerated progress in diagnosis which is to give us increased control over human affairs. But in the practical affairs of the world to-day clinical workers are doing valuable service with makeshift methods of their own devising.

Areas Where the Diagnosis of Conduct Is Needed

The interests of the two groups are fundamentally the same, for both are interested in the development of devices which will help solve certain of the more pressing problems in the control of human relations. These problems may be listed under the headings of *crime, insanity, vocation, and citizenship*.

Diagnosis of Crime. If society is ever to solve the problem of crime, it must have methods of studying the criminal. The diagnosis of crime and of the criminal may have several meanings.

1. *Diagnosis of incipient tendencies toward crime.* Although scientific proof is lacking, there is strong presumptive evidence that most crime is not a spontaneous, accidental occurrence—a “sport” in human behavior—but is usually preceded by peculiar behavior characteristics, which are its direct antecedents extending as far back as early childhood. One may, therefore, speak of tests of predelinquency, tests that recognize delinquent or criminal behavior while in the incipient stages before it assumes serious or tragic proportions.

2. *Prognosis of crime.* Tests prognostic of crime would not differ markedly from the former group. Though it is possible that there may be tests presumptive of a later criminal career which would have little or no relation to present behavior characteristics, the existence of such tests is dubious. Probably the best prediction of later criminal behavior can be had from a diagnosis of present behavior.

3. *Diagnosis of the fact of crime.* This is what is ordinarily meant by the diagnosis of crime. Society usually pays no attention to crime or the criminal until a crime has been committed, whereupon the main question becomes “Did this individual, on whom suspicion rests, commit the crime?”

4. *What commitment shall be made of the individual convicted of a crime* requires for its answer a certain type of diagnosis. The law has one answer to give; medical and social science another answer.

5. *Why did this individual commit the crime?* A study of the factors—environmental and social—antecedent to and causative of the crime should help society determine more rationally what is to be its treatment of the criminal offender.

6. *Prognosis of improvement.* What are the chances that a criminal can be cured? If the theory is believed that crime is a sudden outburst of a fleeting impulse, then it is difficult to make any sort of prognosis; if one believes that the criminal is biologically a degenerated human being, then the prognosis of improvement is hopeless; but if it is believed that most crime is the result of learning determined by environmental influence, then one can estimate the chances that these same habits can be changed, or that these tensions and maladjustments can be corrected.

Diagnosis of Insanity. The problem of the diagnosis of crime and that of the diagnosis of insanity are not entirely separate, because many tendencies which cause a person to become an enemy to society are due to deficiencies in emotional stability or adequacy of adjustment, or definite psychoses or neuroses which are also recognized as symptoms of insanity. We may list several meanings of the diagnosis of insanity as has been done for the diagnosis of crime.

1. In the first place, most psychiatrists recognize that insanity is not a mushroom growth sprouting spontaneously under circumstances of stress and strain. Neurotic or psychotic trends, emotional maladjustments and similar inadequacies of behavior usually have their roots far back in childhood. Where this is the case, these trends are recognizable in childhood or adolescence by one who is familiar with the symptoms. Where these disorders are functional, they should be discovered as early as possible in the child's life in order that some sort of remedial treatment or reëducation can be offered. The discovery of problem children in school and the diagnosis of their ills are very important.

2. Similar to the above is the *prognosis of mental balance and stability*. Probably the same tests, questionnaires, or other techniques which describe the adequacy of childhood adjustments will also be the best obtainable prognosis of future adjustment.

3. Tests are badly needed as *measures of the degree or seriousness of mental unbalance* to be used as criteria for commitment to mental hospitals. At the present time commitment is made on the certificate of a physician with no supporting evidence or tests except the common-sense investigation of the case by the physician and his judgment based on his observations. Although in the long run the interests both of society and of the individual are fairly safeguarded, no one will dispute that such a hit-or-miss

10 Diagnosing Personality and Conduct

method, depending more or less on unstandardized evidence, means that there are wide differences in the decisions. Many are committed as insane who have greater mental stability than others who are unmolested members of society. More definite criteria of insanity and more definite standards by which to judge its degree are needed.

4. We should have not only tests to reveal the fact of insanity but techniques to distinguish the *kind of insanity*. Little progress of definite scientific value has been made in this direction. Insanity is still diagnosed largely by the clinical methods of noting symptoms and making an assignment to one or another of the traditional categories of insanity. As a result of their examinations, two experienced psychiatrists may make entirely different interpretations of the nature of a mental abnormality. Before definite progress can be expected, the present classification of mental disease must be superseded by a classification which is more in harmony with accepted psychological theories. To-day all is hypothesis. Every worker in the field has his own classification, and a classification of mental disorders so fundamental as to command the respect of all has yet to appear. On the diagnosis of the type of insanity, however, depends the decision as to the type of commitment to be made and the treatment to be given.

5. *Prognosis*. After a commitment is made, there is need of a still further diagnosis of the chances of recovery. Here again the methods available are hopelessly inadequate. Methods of dealing with insanity are highly tentative, and consequently little more than a chance prediction of the course of the condition can be given.

6. Intensive research into the *causes of insanity* and into the factors responsible for the increase in mental breakdowns is needed. The adequacy of such investigations will depend on the availability of accurate measures of the individual and his environment, and means for determining quantitatively relationships between the various factors.

Diagnosis of Vocational Competence. In studies of vocational competence most attention has been paid to measures of ability. It is only natural that the first consideration in hiring a worker should be his ability to do the task at hand. But it has been discovered that conduct factors are also important in estimating a worker's value. He must not only be able to do the job, he must

Significance of the Diagnosis of Conduct 11

get along with people and also have certain character qualities of regularity, thoroughness, willingness to stick to the job, and the like. Of fundamental importance is the worker's interest in a given type of work. For instance, he might have the requisite skills and abilities to be a "tree surgeon," but unless he has a definite interest in being out of doors, and in growing things as opposed to working with machines, with people, with books and papers, he would be unhappy and unsuccessful in the work. Vocational counselors are aware that they need measures of interests and of tendencies toward responsibility and regularity to supplement measures of ability, which have already proved their usefulness.

Diagnosis of Citizenship. A fourth area where the diagnosis of conduct is needed is in the work of the schools toward the development of good citizenship. The more tests of intelligence and achievement are constructed, and the better they become, the more painful grows the feeling that certain important educational outcomes are being neglected in the measurement program. Schools are not solely interested in academic achievement and success in subject-matter. Certain character qualities are also important outcomes or "concomitants"* of school work. The development of measures of achievement has tended to over-emphasize the academic and scholarship side of schools, to the neglect of the wider social educational values.

More specifically, education should result not only in the formation of habits, skills, and informations but also in attitudes of acceptance or rejection. In history, for instance, pupils should not only learn certain facts regarding historical events and personages, but they should form certain attitudes toward vital social issues. Tests of historical information have long been available, but measures of social attitude are a more recent development.

The extra-curricular program is one part of the educational scheme for which there has been practically no means of evaluation. With growing recognition of the social values of education and the importance of the social life of the school in the development of habits of good citizenship and favorable personal adjustments, there has come a need for the evaluation of extra-curricular activities. If the extra-curricular program is looking toward the development of qualities of leadership, then methods

* Kilpatrick, W. H., *Foundations of Method* (The Macmillan Company, 1925).

of detecting the development of leadership are needed. If the extra-curricular program aims at the development of qualities of coöperation, good sportsmanship, and social responsibility, then means for noting individual differences in these qualities are needed. There is in this country a National Honor Society for high school students, membership in which is determined by standing in scholarship, leadership, character, and service. Ways of measuring these qualities in high school students are much needed if membership in the society is to have any sort of objective basis.

Types of Evidence

The following forms of evidence are described in this book:

Observation

Rating methods

Questionnaire

to measure adjustment

to measure attitude

to measure interest

Tests

paper and pencil tests of knowledge and judgment

performance tests

The free association method

Physiological measures

Interviewing

Psychoanalysis

External signs

Measures of the environment

Before these are described in detail in the succeeding chapters, there are certain general relationships existing between these different techniques, discussion of which may help in evaluating their significance.

Measures of the Environment, Reactions, or Results

These various types of evidence can be distinguished as descriptions of the *environment*, descriptions of *reactions*, and descriptions of the *results* of conduct. Measures of the environment are important inasmuch as conduct is conditioned by the environment. A rich, stimulating, socially adequate environment

Significance of the Diagnosis of Conduct 13

leads to one kind of conduct; a poverty-stricken, squalid, harsh, or cruel environment leads to another kind of conduct. So important is the environment that it can be used to help evaluate or even predict conduct. Delinquency flourishes in one kind of environment, good citizenship in another. Pupils coming from good homes with thoughtful parental supervision, attending good schools with an atmosphere of social coöperation, and having access to adequate churches and playgrounds are almost certain to exhibit desirable characteristics.

The most obvious way to diagnose conduct is to describe reactions directly, which becomes the second type of diagnosis. The very conduct itself may be on the one hand directly described either by observation and tests, or through judgment by rating methods; or on the other hand it may be obtained indirectly, either through interviews with the individual concerned or with those acquainted with him, or through answers to questionnaires.

Descriptions of the results of conduct have not proven so fruitful as diagnostic techniques. Conduct is supposed to leave its imprint in various physical characteristics such as facial expression, handwriting, posture, and other external signs, but careful experimentation has not revealed these relationships.* Many have thought also that conduct could be diagnosed by studying the products of activity, as for example the results of certain tests, but these methods again have yielded little of significance. Conduct is best studied by catching it in the process.

Direct vs. Indirect Attack

The frontal attack on the description of conduct is often not possible. The observer or examiner is always a factor in the situation, and since conduct is so sensitive to its surroundings, it becomes altered by the presence of the examiner. Because persons are always trying to make a good impression, just as soon as the experimenter announces that he is trying to find out how people behave the behavior itself changes. One cannot announce to a group that he is conducting a test of cheating, for then no matter what else is done, the temptation to cheat is inhibited.

There are times when the direct frontal attack is satisfactory. In cases where there is nothing to hide or conceal, the conduct ob-

* Paterson, D. G., *Physique and Intellect* (The Century Co., 1930).

served may be quite naïve and natural. Persons who enter an interview in order to benefit themselves have every incentive to describe the actual state of affairs. On the other hand, in the involuntary interview one must provisionally accept testimony at its face value, only to compare it for corroboration with other evidence as opportunity occurs. In free association malingering is possible, but with a reasonable amount of coöperation on the part of the subject, the conditions of the experiment almost completely determine the response. What seems to be a perfectly free choice on the part of the subject is really determined by the conditions of his mind. In the questionnaire, however, where the pressure of time is removed, there is a possibility that answers may be given for the effect they produce on the examiner. If the questionnaire permits the individual answering it to show himself to advantage, then there will be every incentive to tell the truth. But if truthful answers reveal things which may be condemned or criticized, then the temptation to falsehood becomes very strong.

In such cases the disguised test or questionnaire becomes imperative. Disguise is quite possible in the case of the questionnaire, inasmuch as apparently harmless or trivial questions often have the power of disclosing significant tendencies when taken with other similar questions. If the questionnaire is headed "Interest Questionnaire" or "Personal Data Sheet" or some other innocuous title, it may soothe any suspicion that it is measuring fundamental habit systems, adjustments, or characteristics of the individual.

Voluntary vs. Involuntary Diagnosis

When a patient goes to his physician for advice and treatment he is in a very different frame of mind from the person who has to fill out his income tax blank or testify in court. The persons who voluntarily seek out an interview with some counselor to obtain guidance in a personal problem are already free from an initial resistance to complete frankness. To be sure, there is an unavoidable human tendency to picture things in as rosy a light as possible, but however much the testimony departs from the facts, it is an unintentional misrepresentation. Those who undertake to be psychoanalyzed have no alternative other than that of revealing the truth. On the other hand, if a person is summoned

Significance of the Diagnosis of Conduct 15

to an interview against his will, either before he is ready to admit his fault or because he is the cause of some other person's distress, then there is an initial resistance and defense that must be broken down. The dean who calls in a student in disciplinary difficulty must first persuade the student that he is friendly, that he is fair, and that he can be of some benefit to the student before the student will reveal himself candidly. There must exist a certain amount of mutual understanding and coöperation between an investigator and his group of subjects before they will answer a psychological questionnaire truthfully.

Standardization vs. Versatility

Another issue in the diagnosis of conduct concerns the matter of standardization. The practical social worker or counselor is somewhat impatient with attempts to standardize diagnostic techniques. To be sure, he will use the Binet Intelligence Test, welcoming the accuracy and impersonality which its standardized technique affords, but in probing into the recesses of individual conduct or into complexes, inhibitions, and obsessions, he may become somewhat impatient with tendencies toward standardization, and skeptical of their value. In the interview there is a tendency to allow the freest sort of intercourse between subject and interviewer. While the subject is urged to express himself openly, to air all his grievances, to make complaints without restraint, the interviewer demands the right to follow up clues wherever they lead. He claims that a prepared schedule of questions can never anticipate the detail that will be necessary to complete the personal story at vital points. The interviewer, he thinks, must discover the plot as he goes along. No prepared schedule can possibly be so adaptable as to fit all the exigencies of highly individual situations.

The weakness in this point of view lies in its overconfidence. Psychology has taught us much concerning the fallibility of human observation, perception, inference, and judgment. The interviewer overlooks his disposition to neglect certain points of the evidence presented in favor of other points; his preliminary bias toward the problem in the shape of pet theories, to support which he is constantly on the watch; and his tendency to stop short of a complete survey of the situation. The scientist is aware that these human weaknesses can contaminate even the description of

inanimate phenomena and recognizes the victory that has been won in the use of exact methods of measurement and experimentation. The tendency toward standardization of the diagnosis of conduct is an attempt to take the fallibility out of human judgment, the partiality from observation and the tendency to generalize on inadequate evidence out of conclusions. The psychological questionnaire or case study schedule helps the investigator base his results on the answers to a large number of questions rather than a limited number. They help him survey all corners of a field rather than become satisfied with uncovering the obvious and convenient facts and then stopping short. The questionnaire, particularly, enables one to make use of proved relationships between answers to a question and inferences from these answers, as against the unproved "hunches" that every worker gathers in the course of his experience. Finally the questionnaire, instead of depending for accuracy on the breadth of experience of the investigator, makes accurate and objective comparisons possible between the answers of one individual and those of other individuals or groups.

The Question of Interpretation

In the last section the matter of interpretation was mentioned as one of the main difficulties that has to be contended with in the diagnosis of conduct. This has not received the full consideration which it deserves. The practical worker in the field tends to discount altogether any suggestion that his interpretation of his observations and inquiries is open to criticism. Perhaps a cocksure attitude is necessary for any one who is going to be influential in practical affairs where the situation is usually pressing, and where the tendency is to give one's best judgment and advice and then back it up with the force of positive authority. Psychological experiments, however, indicate that grave and serious errors lurk around every judgment and inference. These will be treated more at length in subsequent chapters. Suffice it to say that previous experience and our habitual modes of thought tend to warp every judgment made more or less seriously. Genuine progress can be made only on the foundation of patient investigation into relationships, and practical diagnosis can be made accurate only in proportion as the investigator is willing to be thor-

ough and exhaustive and to defer his private judgment to the tested results of experimental inquiry.

In the diagnosis of the insane, of delinquents, and of problem children in courts and mental clinics it is customary to employ both psychiatrists and psychologists. The former have authority and prestige and are better paid; the activities of the latter are narrowly restricted to giving mental tests and recording observations of behavior and characteristics of the patient during the test. The techniques employed by the psychologist are thoroughly standardized, and his reports are highly objective. On the other hand the psychiatrist, in whose hands the disposition of the case finally rests, uses the most subjective of methods, even where certain neurological and medical tests are employed. The psychiatric examination proper is nothing more than a loosely conducted interview whose direction is determined as it proceeds. In contrast to the psychological examination, in which standardized methods are used, the inquiry into the patient's wishes and desires, his thwartings, modes of compensation, moods, obsessions, and the like, is highly subjective. The conclusions reached by the psychiatrist depend for their value upon his fairness, thoroughness, and experience.

Verbal vs. Non-verbal Diagnosis

Before language has developed, the only way in which conduct can be studied is by observing it directly. This method of observation retains its usefulness as a fundamental technique for the study of all ages. But observation is time-consuming and laborious, and therefore as language develops in the child, it becomes a temptation to ask the child about himself. Later, in the study of adults, it is natural to use a person's testimony concerning his own behavior—indeed, it would be unnatural not to do so. This substitution of testimony for direct observation carries with it certain implications which should be clearly understood. To ask a child about his interests, his likes and dislikes, his habits and skills, is to ask him about things that he perhaps has never observed. The greater part of a person's conduct originates without the aid of his verbal organization, and for this reason a special act of learning is necessary before a person steps out to observe himself objectively. To the adult who has grown accustomed to

considering his own problems it seems only natural that the child who has the tool of language at his command can use it with personal reference. However, this assumption needs to be tested.

Along with the question of the possibility of self-observation goes the matter of the voluntary alteration of language to fit purposes. It is relatively difficult to change conduct itself in the presence of the examiner, for though one can act hypocritically for a short time, sooner or later the true response will reveal itself. But language is ever so much more versatile than conduct. There need be only the loosest connection between language and conduct, between what a person says he will do and what he actually does, between what he says he believes or likes and what he actually believes or likes as shown by his conduct.

Notwithstanding this thread-like relationship, the results of questionnaires and interviews have demonstrated their usefulness. But the empirical approach which is referred to again and again throughout the pages of this book must be emphasized here. Answers to a questionnaire must be correlated with other known characteristics of individuals, and the answers must then be interpreted in the light of these correlations. Answers must not necessarily be taken at face value, for their implications are often subtle and obscure. The findings of statistical inquiry often reveal significances to questions and answers not dreamed of by the investigator. Verbal evidence is to be valued highly in the diagnosis of conduct. It is of all forms of evidence the easiest to obtain, to record, and to study objectively. But its significance must be tested and discovered.

Honesty

At several points in the previous discussion the question of honesty arose. In the indirect approach to conduct diagnosis, the question of honesty does not enter, but where direct questions are asked a subject and the diagnosis depends on the truthfulness with which he answers, honesty looms large as an issue. Evidence is slowly accumulating concerning the honesty with which children testify concerning themselves. It is by no means 100 per cent. Sometimes the error may be ascribed to ignorance, sometimes to purposeful deception based on a definite desire to conceal unpleasant truths. Honest replies depend in part on the voluntary nature of the testimony; and in part on the degree to

Significance of the Diagnosis of Conduct 19

which the subject believes his testimony will be approved. Instead of condemning questionnaires in a wholesale fashion because they may yield dishonest evidence, we should assume the attitude that here is a disturbing element which must be studied and its influence discovered under varying conditions so that it may be controlled or allowed for.

Standards of Conduct

When one commences to investigate conduct he enters the field of morals, and as soon as an issue becomes a moral one, the question arises: By what standards do you judge conduct?

Insofar as possible in this book, the matter of standards is avoided. The main emphasis in diagnosis should be upon *description* rather than *judgment*, upon the endeavor to describe conduct fully and accurately without trying to evaluate it or pass judgment upon it. In practical affairs, however, standards of judgment and evaluation are necessary. Society sets up rules of conduct to guide its affairs, and these become embedded unconsciously in its *mores*, or embodied consciously in its laws. Every institution—the school and home particularly—sets up standards of conduct to guide its members. In this book these standards are accepted as they have been formulated by students or lawgivers. If they are standards of conformity, then there are methods of rating and observation for studying these facts! If they are standards of adjustment or integration or efficiency or self-expression, then there are various techniques for discovering these facts. In measurement the scale defines the standard, and the procedure then becomes one of merely stating the facts in terms of more or less on the scale. This book is concerned less with what to measure than with how to measure.

Occasionally, when one must make some sort of evaluation of the items of a test or questionnaire before it can be used to measure, experts or qualified persons are called upon to decide as to the direction or tendency of various answers to questions. If one is studying knowledge of health conduct, then experts must decide what answers represent acceptable health standards. If one is using questions to measure the liberalism-conservatism of a group, then the questions must be evaluated by competent persons who decide which answers lean in the liberal direction and which

Diagnosing Personality and Conduct

in the conservative. There are also empirical methods of determining the direction tendency of answers to questions by tabulating the responses of different groups already socially selected according to the standards which one is measuring. For instance, the different ways in which delinquent and non-delinquent groups answer various questions enable one to determine the significance of these questions as measures of delinquent tendencies.

In the case of purely moral issues where standards are determined by the expert judgment of some group, it should be kept in mind that moral standards are purely relative, and that no group can claim authority in the pronouncement of moral standards. Since Hartshorne and May, in order to construct a scoring key, referred their tests of knowledge of right and wrong to a group of educators interested in character education, the tests represent nothing more than the standards held by this one section of the community.

Standards of conduct are most trustworthy when based on objective evidence that they satisfy certain accepted values. For instance, if *health* is an accepted value, then scientific investigations which yield evidence of the healthful qualities of activities certify to the acceptability of these activities. Likewise if *efficiency* is an accepted value, efficient behavior can be found by experimentation. To discuss the question of standards adequately would require a volume in itself. In this book *methods* of diagnosing conduct are discussed. The other issue of validating this conduct must be left for a subsequent treatment.

Reliability of Diagnosis

A primary concern of this book is the accuracy or reliability of diagnosis or measurement. In every chapter pains have been taken to comb carefully the technical literature in order to discover and present the evidence bearing on reliability.

The significance of the coefficients of correlation which show the agreement between repetitions of the measures of the same phenomena is discussed in books on statistical methods. There are two types of evidence on reliability. One is the correlation of one half of the measuring instrument against the other half (the halves chosen at random) giving a measure of the *consistency* of the instrument. The other is the correlation of the measurement with

a repetition of the same measurement on the same individuals at a later date. This involves not only the consistency of the instrument but also and primarily the consistency of performance of those being measured. In measures of ability the latter is the more important because we want to know not only how consistent from part to part the instrument is but how reliable a single application of the test is for measuring an individual under all times and circumstances. In the diagnosis of conduct, however, one does not expect nearly as much consistency of individual performance as he expects in the measurement of ability. Where abilities change only very slowly, conduct may shift rapidly with the changing situation. Hollingworth found that neurotic World War soldiers, tested immediately after the armistice, made responses to a questionnaire quite different from those they had made before.

But one is concerned primarily with the consistency of the instruments themselves, for he is never quite sure of the real existence of many of the categories by which human nature is described until these have been tested. For instance, it is claimed that answers to certain questions are symptomatic of introversion or extroversion. If a list of these questions is assembled and given to a group of people, this assumption can be tested. If the questions are sorted into two groups in a chance manner, and individuals who answer the questions in one group with a decided tendency toward introversion, answer the questions in the other group with no tendency at all, much doubt is thrown on the ability of the questionnaire to show the introversion-extroversion difference. If on the other hand, persons showing strong introversion tendencies on one half of the questions also show these same tendencies on the other half, the instrument operates with high consistency or reliability. The coefficient of correlation between the scores on the two halves measures this tendency toward reliability. The reliability of a diagnostic technique is one of the *most important* facts we can know about it. If it is unreliable, it is not measuring anything accurately and is useless in its present form. However, if the technique has promise of being significant, the unreliability can perhaps be remedied by extending the sampling or standardizing the procedure.

Do we "traumatize" human experience by our methods of measurement? Do we assume from a method of measurement whereby we ask a number of questions and then proceed to total the an-

swers so as to yield a total score that there is something in human nature that corresponds to this score? It must be admitted that this is oftentimes the case. It is easy to fall into the trap of first positing a trait, such as persistence or dependability, analyzing the trait into a number of specific situations in which it might manifest itself, framing these in question form, and finally computing a score on the basis of answers which indicate the presence of the trait. This method makes the unwarranted assumption first of all that the trait actually does exist.

One must be constantly on his guard in measurement not to create falsely by the very names which he gives to his tests a fictitious relevance or significance which they do not actually have. Psychology has suffered grievously in the past from an exuberance of imagination on the part of its builders which has led them to create concepts with no corresponding reality as proved by their demolition by later experimentation. *Hypostatization* is the name the logicians give to this fallacy. The whole *faculty* psychology was erected by assigning a reality, concreteness, and independence to certain concepts of mental life that subsequent investigation shows do not have a corresponding independent existence. Likewise the present *trait* psychology is under the very strongest suspicion of being similarly unfounded. This should put every investigator on his guard against creating psychological fiction by the very names which he assigns to his tests.

The check for detecting this fallacy is measurement of the reliability of the tests, questionnaires, or other instruments used. If the instrument has high reliability, this is evidence that there is something there in reality, even if the name assigned to the test may not properly describe it. One may speak of a questionnaire to measure introversion-extroversion when in reality it may merely measure the subject's tendency to think of himself as introverted or extroverted. What the instrument really measures must be determined by the correlations of the instrument with other factors. But the reality of the *confact** must be determined by the high reliability of the instrument.

* Symonds, P. M., *The Nature of Conduct* (The Macmillan Company, 1928).

Chapter II

OBSERVATION

AS a technique of investigation, observation is of prime importance. All experience, all data, must enter through the portal of the sense organs. There is nothing that can be known or studied that does not have first to come under observation. This truism is often overlooked because sometimes we make our data so accessible, so easy to read and interpret, that the problems of observation reduce almost to the vanishing point. For instance, when we reduce the measurement of weight first to the simple manipulation of scales so that they come to a balance and then to the reading off of a figure marking the balancing point on the measuring scale, errors due to observation are negligible. On the other hand, if we have to judge weight by lifting another person in our arms, errors of observation are considerable. Again, if we reduce a person's ability to a series of written responses in the form of crosses or underlinings to certain standard exercises, and we score these responses by comparison with a previously prepared key, the only errors of observation are those which result from errors in checking the responses against the key or in counting the correct responses. But if we have to depend for our estimate of a person's ability on scattered impressions based on casual observation of his behavior in whatever situation we can find him, our judgment will be extremely liable to error.

Advance in scientific method has consisted in large part in devising instruments and techniques that help reduce the errors of observation. Sometimes phenomena are magnified so that they can be more easily grasped by the senses, permitting smaller differences to be observed, as with the telescope and microscope. Sometimes differences are brought closer together in space and time so that comparisons and discriminations can be more accurately made. The most efficient method of eliminating errors of observation is to reduce the phenomena to a permanent form to be studied at one's leisure and convenience, so that errors due

24 Diagnosing Personality and Conduct

to observation "on the run" and to faulty memory may be avoided.

Many of the techniques described in this book are fairly objective, so that errors of observation play only a minor part in them. But the progress of the science of measurement of conduct has not proceeded far enough to provide techniques for the objective measurement of most of the phenomena of conduct. We must still rely largely on direct observation of conduct itself, employing all of the cautions and safeguards and improvements that have grown up in the technique of observation.

In studying the behavior of infants, investigators have had to depend principally on observation. At this early age articulate responses are not available, and hence cannot be used as a source of evidence. Consequently most of the advance in the technique of observation hails from the nursery schools and child development institutes. Valuable contributions have been made in this field by Olson (8)* and D. S. Thomas (11). But the value of direct observation is not limited to the early ages, and indeed the techniques to be described might well be borrowed and applied to the study of all ages. Direct observation is a fundamental source of data which is at the same time a welcome complement to, and a check upon, the sometimes too heavy reliance on tests and questionnaires.

In this chapter we shall first discuss the fundamental conditions of observing and perceiving and analyze the errors that frequently occur. Next the question of selecting and defining what is to be observed will be discussed. This will lead to a treatment of sampling, units of measurement, and reliability in observation. Finally, certain practical questions concerning observation and methods of keeping records will be considered.

Essentials to Observation

The first prerequisite to accurate observation is the possession of efficient sense organs. In practice it turns out that most observation is visual. The eyes are said to be the most sensitive of all of our sense organs, and a good pair of eyes with lenses that focus, both at near and far distances, and without astigmatism,

*Numbers in parentheses refer in each case to references at the end of the chapter.

is an important adjunct for successful observation. Glasses may correct certain defects, while at the same time adding difficulties. Color-blindness is a defect of the eye which renders complete if not accurate observation impossible. Of course not all sense data may come by sight, and hence acute and sensitive sense organs for hearing, taste, smell, touch, pressure, and temperature, not to mention the kinæsthetic senses, are necessary for complete observation.

A second essential to good observation may be summed up under the term *alertness*. To make a correct observation one must attend to the object being studied. To direct attention to the right place is not the simple and obvious matter that it may seem. In obedience to the laws of attention, competing stimuli often attract the attention to the wrong place or at the wrong time. Psychologists tell us that we attend to changing stimuli; that, other things being equal, the strength of the stimulus is a prepotent factor; that repetition of the stimulus dulls the attention; and that striking quality or definite form attracts. Working against these, competing interests or competing stimuli distract.

The fact that the attention is drawn powerfully and compellingly to certain stimuli rather than others is made use of by every magician. In stage tricks the magician usually manages to divert the attention by his own movements and the direction of his own gaze while an occupied hand is shifting the cards or putting the rabbit in the hat. Münsterberg (6, p. 29) demonstrated this experimentally on the lecture platform of his classroom as follows:

"I stood on the platform behind a low desk and begged the men to watch and to describe everything which I was going to do from one given signal to another. As soon as the signal was given, I lifted with my right hand a little revolving wheel with a colour-disk and made it run and change its color, and all the time, while I kept the little instrument at the height of my head, I turned my eyes eagerly toward it. While this was going on, up to the closing signal, I took with my left hand, at first, a pencil from my vest-pocket and wrote something at the desk; then I took my watch out and laid it on the table; then I took a silver cigarette-box from my pocket, opened it, took a cigarette out of it, closed it with a loud click, and returned it to my pocket; and then came the ending signal. The results showed that eighteen of the hundred had not noticed anything of all that I was doing with my left hand. Pencil and watch and cigarettes had simply

not existed for them. The mere fact that I myself seemed to give all my attention to the colour-wheel had evidently inhibited in them the impressions of the other side. Yet I had made my movements of the left arm so ostentatiously, and I had beforehand so earnestly insisted that they ought to watch every single movement, that I hardly expected to make any one overlook the larger part of my actions."

In the hurly-burly of a complex situation it takes considerable skill to direct the attention properly to the point of critical issue. Many are unable to enjoy a football game because they do not know how to observe. "Beautiful interference," "a hole between center and left guard," "an incompleted pass" are mysterious terms because the attention is in the wrong place. So in a classroom one may not direct his attention properly to discover how pupils react to the complex social situation. In describing her experimental work with nursery children Miss Thomas (11, p. 6) notes,

"The ordinary observational technique was shown to be totally invalid as an instrument for recording behavior for research purposes. A check-up was made by having several observers make diary records of the social behavior of given children. There was a marked tendency for a given recorder to note one aspect in the record; another in the next record. Although trained to make objective records in the sense of including overt reactions only, the result was necessarily subjective in that the recorder, usually unconsciously, selected specific parts of the total behavior act. This selection became, furthermore, invariably inconsistent over a period of time. The obvious solution to this difficulty was to break up the behavior-complex into relatively simple units which would enable a record to be made of every recurrence of one of these behavior units."

A third essential to good observation in some situations is the ability to make reasonably accurate estimates without the use of special instruments. When they are available, measuring instruments help us to dispense with some of the skill necessary for discrimination. This ability to make accurate estimates is something that may be improved by training. Indeed, in much of their casual experience young children are increasing their ability to make accurate estimates by observation. They must learn to estimate distances and to gage their movements accordingly, learn how far to reach out a hand, or how long a step to take. But these

estimates, accurate enough for the practical business of living, especially since an error is easily corrected by a new adjustment of reaching a little farther or taking another step, are woefully inaccurate when we must make a record of our impressions, especially initial impressions. Similarly, judgment or estimates of number, height, weight, volume, intervals of time, brightness, color, pitch, volume of sound, temperature, taste, smell, and the like may be and usually are extremely inaccurate. Such estimates based on observation are shown clearly to be in error when comparisons are made between them and objective records.

The ability to make accurate observation is something that can be increased by practice up to the point, indeed, where some persons by long experience and directed practice have gained remarkable skill in accurate observation. Winch showed how the ability of school-children to make random observations can be increased measurably. Every one, by devoting attention to the matter and practising, can become highly accurate in ability to estimate distances, lengths, weights, volumes, etc.

A fourth essential in observation, the capacity to make fine distinctions, grows directly out of the previous point. Accuracy in estimating leads to accuracy in making comparisons. One who is able to estimate accurately the length of a line is also thereby able to judge more accurately which of two lines is the longer. The mother is alert to small changes in her son's behavior; she recognizes even slight irritability as a possible symptom of illness or fatigue. The interviewer learns to detect slight indications of emotion or nervousness in an individual and uses these signs to help direct his questions to bring out significant items of information. Just as some persons at a concert appreciate every shade of interpretation of the artist or cringe at technical errors, while others applaud the good and the bad equally, so certain of us remain obtuse to the social forces at play in a situation, while others are sensitive to every nuance.

A fifth essential to accurate observation is freedom from various pathological states. A fatigued person cannot observe accurately; his attention lapses, and things pass by unheeded. Alcohol and drugs have well-known effects on the power to observe. They lessen accuracy and impair the balance that is necessary for correct weighting of the various factors involved in correct inter-

28 Diagnosing Personality and Conduct

pretations. Feeble-mindedness or insanity renders a person incompetent to make accurate observation.

Still another essential to observation, the sixth in our list, lies in *making an immediate and accurate record*. It is well known that memory fades. The clear-cut perception which one may have at the moment dims with the passage of time, and the memory image becomes indistinct. Testimony in court is often inaccurate, not because the original observation was faulty, but because in the interval between the event and the recall the memory of the event has grown dim. Any one who wishes to preserve the results of accurate observation should plan to make an immediate record of them.

A seventh essential to good observation lies in the ability to *perceive* accurately, for when that stage of observation which is called perception is reached, the opportunity for error increases tremendously. In perception we make an interpretation as to the significance of what our sensory data includes. What we see or hear is taken as a cue of the presence of a whole object or situation of which the stimulus is only a part. This tendency to generalize on the basis of a limited experience is utilized by the radio, which employs such devices as the clapping of two sticks together to indicate a revolver report, or the roll of a drum to indicate distant thunder. These interpretations of sensory stimuli are largely a matter of habit.

Now the observation of human conduct is full of such interpretations which are at the same time necessary and liable to error and unreliability. The doctor diagnoses a disease on the basis of a few symptoms; the psychiatrist diagnoses a neurosis or a family situation on the basis of a few questions and observations. But the opportunities for illusion are numerous. The child buried in a book may not be studying, but day-dreaming; the awkward, clumsy girl may have been growing so rapidly as to outstrip her powers of muscular coordination, or she may be trying to attract attention. We tend to credit the child with pleasant expression or a sparkling eye, or the child who wears glasses, with intelligence. We fail to look for the brightest child in a class-room among the youngest and smallest. The boy who gains our applause by some cheap trick of foolery is really covering over some other weakness or deficiency from which he wishes to distract our attention. Heinlein (4) found that observation of children's rhythm in

marching was much disturbed by the rhythm in the music which accompanied the marching. To observe correctly is one thing—to interpret what we see is a different matter.

Miss Thomas (11, p. 9) notes inaccuracy in perception in her experimental work as follows:

"One reason for our low reliability on the *number* of social contacts (irrespective of timing, the Pearson *r*'s ranged from only .47 to .80 for the six observers in the preliminary reliability study) is our neglect to differentiate between the merely spatial and functional social relationships. Hence, when a given child approached another, the recorder made his own interpretation of whether it was a genuine social contact. These low coefficients of correlation are a beautiful example of how unreliability creeps in where interpretation is permitted the recorder."

To state as an eighth essential of good observation *freedom from prejudice or from habits of interpretation* is perhaps merely to reinforce the value of correct perception. But to place stress on the habitual nature of our perceptions is important. We grow up to see things from the point of view of the use we make of them, our relations to them, or we make analogies in interpreting new phenomena in terms of our response to the familiar. The hen responds to the shiny oval piece of porcelain in her nest exactly as to a real egg, and lays another beside it. Hunters use decoys to attract their game or feign the cries of the animals which they wish to shoot. It is most difficult for most persons to describe without interpreting, yet this is the essential of good observation. The teacher brands a pupil as lazy, but unjustly. What she should do is to observe and describe his behavior as well as its antecedents and then survey the possible interpretations—illness, bad habits, competing interests, inability to cope with the task, etc.—before deciding on one. Every supervisor who observes a class-room lesson enters the situation full of prejudicial habits of thinking. He may be a supervisor of the old school who is horrified by the least sign of disorder, or a supervisor of the new school who is shocked by regular rows of desks and a formal question-and-answer recitation. So completely is observation warped by interpretation that two visitors to the same class may make quite conflicting reports. If supervisors could be taught to *observe* teachers, or teachers taught to *observe* children, accurately and precisely, and then make their interpretations, judgment of con-

duct would be considerably improved. The social scientist who would use observation as a method of investigation must approach the situation to be observed free from indoctrination, ready to describe what is there without bias.

Clarke (2) in *The Art of Straight Thinking* points out what we shall list as a ninth essential to good observation: *freedom from excitement*. When the emotions are aroused, all conduct becomes impetuous, impulsive. The tendency to *interpret* instead of *observe* becomes reinforced. Mere observation with postponement of interpretation becomes impossible under excitement. We demand action, and action requires interpretation of the situation. The moral of this is to refuse to make observations when a crisis arises. Do not wait until a child commits a misdemeanor before starting to diagnose the situation. It is hard for a person under suspicion to escape being misinterpreted. Every misstep, every slip, every aberration is seen as additional corroboratory evidence. Observation should be made systematically when there is no crisis, and its results should be recorded to be used when the crisis arises if necessary.

Selecting and Defining What Is to Be Observed

Observation may be divided into two types—*finding observation* and *directed observation*. In the former one knows simply that he is to observe a given situation, not what he is to look for. It may be that he is to try to discover the factors inherent in the given situation, as when the writer (10) once conducted a series of observations on the study activities of high school boys. He admitted before he began that he did not know what study was, or how it was constituted, but he could state that his purpose was to make a survey of the activities constituting study by observing boys while they studied. To conduct such a *finding* observation the observer must be generally competent in the field and must know the possibilities. Every finding observation is restricted by the limitations of the observer's background of experience. For example, in attempting to discover the factors entering into high school study, I might have neglected various *clerical skills* of copying, using the alphabetical index, and the like, if I had not been sensitive to these factors from my general experience with various sorts of reactions.

Directed observation on the other hand is quite definitely limited to the schedule used. A list of acts which has been previously prepared provides the basis of the observation. Attention is concentrated on the occurrence of the items in the list, while everything outside of this list is ignored.

The *finding observation* must precede the *directed observation*, because only on the basis of a finding observation can the inventory or check list to guide the directed observation be constructed. Andrus, (1) in her preliminary work in inventorying the habits of young children, had her observers keep *diary* records which were merely running notes of every act of the child under observation. Entries in these diaries were then analyzed into separate *habit* lists, and from these the inventory list was constructed. Enough diaries were used to ensure that the check-list would not be appreciably increased by the use of additional diary records. This check list then became a guide to be used in subsequent directed observations.

Other more analytical methods of selecting the check list to be used in directed observation have been employed. Olson, for instance, derived an inventory of ties and nervous habits in children by a tabulation of their mention in the literature on nervous and mental diseases. This method, of course, depends on the observation of other persons who are recognized experts in the field. After Olson (8) had completed the first part of his investigation, he determined the correlation of each of the five groups of nervous habits—oral, nasal, hirsutal, ocular, and aural—with the total number of nervous habits, and found that the oral habits gave the highest correlation, $+ .77$. This was taken to indicate that oral nervous habits are the most symptomatic group, and so in later work he concentrated his attention on this particular group. This empirical method of selecting items to be observed has much to commend it.

Miss Thomas states that she selects her items for observation both on the basis of their *social importance* (validity) and also on the reliability with which they can be observed and the degree to which they can be observed with freedom from prejudice. The latter is mainly a matter of trial, but also depends partly on the way in which the items are defined and limited. For determining the importance of the items she has recourse to finding observations:

32 Diagnosing Personality and Conduct

"Not only must we emphasize *how* we obtain data, but what we obtain must of itself be important. It may be questioned as to how we know we are getting at what I have called 'important' aspects of social behavior. We say 'important' advisedly—for some selection on an *a priori* basis is necessary in statistical investigations. We consider correlation coefficients and other instruments depending on the theory of probability as poor means of discovery without a preceding intimate knowledge of the way in which the data behave, but excellent instruments for controlling bias and evaluating relationships if we first know our materials." (11, p. 19)

Olson, however, made a perfectly justifiable use of correlation in selecting important items.

An adequate description of the items to be observed is very important, for the precision with which the items are described determines the amount of interpretation which the observer needs to make. Miss Loomis (11, ch. III), working under Miss Thomas, found that there was little difficulty in recording the number of physical contacts one child made with another child, but that there was difficulty in classifying them into such categories as "hit, point, pull, push, caress, explanation, accident, or assistance." Of two observers making simultaneous records on the same situation, one recorded, "Conflict between William and Edward," whereas the other observer recorded it "William embraces Edward. . . ." (11, p. 11) To reduce the subjectivity of interpretation is the purpose of detailed descriptions of the items to be observed.

Observation as Measurement

Once the items for observation have been selected and defined, one must decide upon the unit of measurement. Newcomb (7) in his study of problem boys carried a pack of cards around in his pocket and simply jotted down incidents as he noticed them as they occurred at one time or another during a summer camp period. The unit of measurement here was the occurrence or non-occurrence of a given act over a period of time. Others who have more control over the situation have tried counting the number of occurrences of an act in an interval of time. But as Olson found in connection with his nervous habits, sometimes a child will continue to keep his thumb in his mouth continuously while another child will keep putting his thumb into his mouth and taking it out.

In such a case a count of the "number of times" a child puts his thumb in his mouth is of little value.

An advance over these methods of measuring in observation is to break the period of observation up into short intervals and note the occurrence or non-occurrence of the act or habit during each interval. These intervals may be mere segments of a continuous period, or they may be intervals coming on separate days. The latter method will give the better sampling. Olson tried observation periods of five and ten minutes, one a day for from ten to twenty days. His technique of using as the unit of measurement the occurrence or non-occurrence of an item within a short time interval is a distinct contribution to the method of observation as a tool for scientific experimentation.

Miss Parten, working under Miss Goodenough (3), extended Olson's technique by obtaining forty observations over *one-minute* intervals. The empirical results were also almost entirely in agreement with the theoretical predictions from twenty observations by the Spearman-Brown formula, indicating that almost any desired reliability can be obtained by increasing the number of observational samples. Later ten-second intervals were used, a check being made at the end of each interval of the presence of the act or behavior under observation. Seemingly there is no limit to the increase in reliability to be obtained by (a) decreasing the size of the observation interval, (b) increasing the number of observations, and (c) properly spacing the observations to take care of the sampling of varying concomitants of the situation such as group activity, time of day, fatigue, persons present, etc.

Exactly how these three variables in the measuring process of observation will be weighted must be decided empirically. Several factors are concerned, and some traits may require longer time intervals than others. Simple situations, such as movement vs. passivity, laughing vs. not laughing, reading vs. not reading, need only short time intervals. Other more complex traits, such as cheating, borrowing, teasing, and the like, need longer time intervals. The frequency of occurrence of the behavior is another factor that must be considered in deciding upon the time interval. "Making the assignment" as an item in teacher observation comes only once a class period, and very short observation intervals are out of the question. The number of observations is also determined by the objectivity of the trait to be rated. Traits in which there

is a subjective element need more observation in order to achieve the desired reliability. Constancy of manifestation of the habit is still another factor.

Miss Thomas raises a pertinent issue in this connection. If one is interested in studying growth or change or development, the observations must not be strung out over too long a period of time. Indeed, the aim in this case should be to get as instantaneous a picture or cross-section as possible. Some sort of compromise must be made between extending the observations for the purpose of obtaining reliability and compressing them in order to get a more nearly instantaneous picture. This issue will be particularly important where observation is used to record the result of learning experiments conducted over short periods of time. One solution is to increase the number of simultaneous observations by different observers.

Miss Thomas points out that there are two questions involved in the reliability of observation, one of which is the *consistency* of a given experimenter's observation over a period. This is partly a matter of the skill of the observer, but is largely determined by the sampling and range of items to be observed. Observation in a narrow situation may yield more reliable results than observation when the situation varies. Again, in the variable situation greater reliability can be achieved by extending the observation.

The other type of reliability concerns the *agreement of two observers*. Here the matter of objectivity enters: two observers will agree more closely if the behavior to be observed is objectively defined and if interpretation is not called for. Skill of the observer, sampling of situations, and number of observation periods all enter into the reliability of the observations of two judges. Of the two types of reliability, the agreement of two judges is more important than the consistency of a single judge because the former includes the latter.

Olson finds the reliability in observing oral nervous habits in children to be .76 for ten ten-minute observations and .87 for twenty five-minute observations (240 cases of school children from age eight to age twelve). Similar figures for 169 children in departmental classes in school were .69 (ten ten-minute observations) and .82 (twenty five-minute observations). Quite obviously the advantage in reliability goes to the larger number of observations and the shorter observation interval.

Olson (8, p. 24) also shows that the constancy of oral habits decreases with increasing intervals between observation periods.

TABLE 1
CONSTANCY OF ORAL NERVOUS HABITS OVER AN INTERVAL
(From Olson)

Grade	No. of subjects	Interval in days	Method *	r	r corrected for attenuation **
6A	25	8	1	.80	.98
5B	31	4 ¹ / ₂	2	.50	.57
3A	31	35	2	.51	.59
2B-2A	29	49	1	.49	.60
2B-2A	32	158	3	.26	.32
1A-2B	26	150	3	.49	.59
3A-5A	49	365	2	.40	.46

* Method 1. Ten ten-minute observations vs. twenty five-minute observations.

Method 2. Twenty five-minute observations vs. twenty five-minute observations.

Method 3. Fourteen ten-minute observations vs. twenty five-minute observations.

** In correcting for attenuation the reliability of the ten ten-minute observations was taken as .76, for the fourteen ten-minute observations as .80, and for the twenty five-minute observations as .87.

Although certain inconsistencies in the technique tend to make the correlations over long periods too low, the evidence shows that with accurate observation there is high consistency of habits over short periods, but that this drops off over longer periods. Even after a year, however, there is considerable constancy in the number of oral nervous habits that a child retains.

Olson also gives reliability coefficients showing that some nervous habits are observed less reliably than others, partly due to the difficulties in observation.

Miss Thomas reports several sets of reliability coefficients for observing different functions. In a study by Miss Barker in which the time spent by children in different activities was observed, the reliabilities ran from .92 to .96. This, however, does not include errors of the type in which one observer fails to check an activity noted by the other observer, a type which was present in a considerable proportion of the cases. When observers are trained to time *all* activities, the reliability rises to .95-.98. Similar reliabilities (.97 and .98) were found in the observation of two observers concerning the amount of space covered by children in

their spontaneous activity. Lower reliabilities (.47 to .80) were found for "number of social contacts," but this is due to lack of agreement as to how *social contact* shall be defined. The reliability for observations of two observers on percentage of *time* spent in social situations by each child was .69 and .86; for the *number* of *social situations* recorded for each child, .85 and .86; and for the *number of children* played with by each child, .77 and .88.

Miss Goodenough reports the following reliability coefficients on observations by two independent observers, whose combined observations did not occupy, on the average, more than twenty minutes for each child, and which were not made simultaneously. It is inferred that the time interval taken as the unit of measurement is ten seconds.

TABLE 2
RELIABILITY COEFFICIENTS FOR OBSERVATION
(from Goodenough 3, p. 234)

Physical activity58	Leadership74
General activity80	Anger71
Laughter32	Dramatic play87
Conversation72	Self-help35
Social participation60	Response to food.....	.50
Reluctance.....	.71		

These reliabilities are not high, but show what can be achieved with untrained observers and under somewhat unfavorable circumstances.

Conditions of Observation

The necessity that the observer be in a favorable position to observe is so obvious a factor that it needs little discussion. Olson, for instance, took a position at the front and side of the room where he could observe the faces of all the children in the class. Much that passes for observation, however, is obtained from only fleeting glimpses, with obstructions to the view, and in face of distractions of various kinds.

One critical issue in the technique of observation is the amount of control of the situation which should be exercised. On the one hand, we may make observation in a strictly laboratory situation with various factors such as light, temperature, and material and social stimuli rigidly controlled. On the other hand, we may wish to exert no control whatever over the situation. In the laboratory

type of situation the number of variables is reduced so that the naked relationship between two variables is more clearly apprehended. On the other hand, there are those who maintain that the most valuable information comes from judging how a child reacts in a normal social environment as it happens to develop. The question at issue here goes back to the fundamental nature of social behavior. Some maintain that the factors operating are so complex that it is not possible to take Humpty-Dumpty apart in the laboratory and study the pieces, and then properly know how these pieces will work when put together again. That the whole is not equaled by the sum of its parts is the belief of many students of social reactions.

Certainly greater control could be obtained if it were possible to study the independent influence of separate factors, from which a certain amount of generalization would be possible. If we are reduced to describing the behavior of total situations, the number of different situations to be investigated becomes discouragingly large. But it must be admitted that the problem is not so simple as in the physical sciences, where the influence of each factor is observable. In the social sciences the factors are more numerous, seemingly defying control and measurement, and hence the immediate advantage of analysis is not visible.

A kindred issue concerns the part the observer should play. On the one hand there are those who would remove the examiner entirely from the situation. If he is present, it is agreed he should be a passive and unobtrusive visitor, but it would be better still if he could be invisible. The Yale Psycho-Clinic has devised a screen behind which the observer becomes invisible to children in the room, and is therefore eliminated as a stimulating factor in the situation. The writer became skilful in observing the study habits of boys through a technique of concentrating on a certain boy until the boy looked at him, whereupon he would turn his attention elsewhere. It is possible to disarm suspicion by apparently being unaware of what is going on. At the other extreme is the observer who actually takes an active part in the situation by asking questions, making suggestions, teaching, and the like. Observation may be an important feature of an interview. The part that an observer takes in the situation depends on the particular problem under investigation.

Quite apart from its bearing on the matter of reliability, there

38 Diagnosing Personality and Conduct

are issues connected with the length of the observation period. Sometimes, as in the study of sleep or the influence of drugs, the observation may be carried over extended periods. Again, if it is a case of merely studying the frequency of certain simple behavior items, the period of observation may be very short. To observe a teacher for fifteen minutes in order to formulate a judgment as to that teacher's efficiency is ridiculous. A complex process such as teaching cannot possibly be adequately sampled in so short a period. On the other hand, Olson found 100 minutes to be adequate for the reliable observation of nervous habits of a class-room of children.

When one is studying growth or development, the observation may be carried on for months or years, although of course not continuously. Certain writers have kept diaries and others accurate records of their own children's development based on observation carried on over long periods of time. On the other hand, an instantaneous cross-section picture may be what is desired.

Whether to concentrate the observation on one child or on several children is another issue. Naturally it would be more economical if more than one child could be observed at a time. In general one may say that in a finding observation the attention should be devoted to one child or to a very small group. If, on the other hand, the observation is directed by a very short schedule of items which are easily observable, more than one pupil can be observed at a time. The writer in his investigation of study habits followed one boy around the school for a whole morning. His purpose in concentrating on a single boy was to make an intensive study of everything that one boy would do relating to study procedures. If, on the other hand, certain items of study behavior were listed for observation, it might be possible to cover the activities of a whole class-room or study hall at once by the use of seating charts and symbols to represent different activities.

Recording Observation

Finally the importance of record-keeping in observation must be emphasized. The development of knowledge and the scientific method depends on the accumulation of records. Observation is resorted to only because the phenomena to be studied are fleeting and transitory and do not leave their own records. We cannot

depend on memory, because impressions are nearly as fleeting as the phenomena themselves. Items of observation not only become lost, but are mutilated or altered when the memory alone is charged with their preservation. Observations should be recorded on the spot, since impressions lose their vividness even after a few hours.

In *finding observation* the records must be in the nature of running commentaries or diaries. While investigating study habits the writer kept at hand a small note-book in which every act of the boy under observation was stealthily entered. In observing small children the record need not be kept out of sight, as its meaning is not understood and consequently it does not become a disturbing factor in the situation. These diary records should be revised for permanent filing because in their original state they are apt to be fragmentary, sketchy, and full of cryptic symbols.

In directed observation a check list is used, and the technique of recording is correspondingly radically altered. In the first place, the observer must be thoroughly familiar with the inventory or check list. There is usually no time to run over the list in search of a particular item. Its place in the list must be known ahead of time. Usually it is possible to prepare special sheets so that the occurrence of an item can be checked. A favorite device available in class-room observation is the seating plan, in which each child's name is inserted. Armed with such a seating plan the observer can record through a set of symbols the events of the class-room period, the identity of those asking questions, the identity of those answering questions, and the nature of the responses. Miss Barker (11, ch. II), in studying the movements of nursery school children, made use of a floor plan of the space on the roof of the building where the children engaged in free activity. By the use of lines to denote a child's movement and position, and symbols to denote the nature of the activity, certain aspects of behavior were recorded so as to be susceptible to accurate measurement. The use of these diagrammatic records of observation, be it noted, requires a certain amount of skill and practice.

REFERENCES

1. ANDRUS, R., *Tentative Inventory of the Habits of Children from Two to Four Years of Age*, Teachers College Contributions to Education, No. 160 (1924).

2. CLARKE, E. L., *The Art of Straight Thinking*, (D. Appleton and Company, 1929), ch. V.
3. GOODENOUGH, F., "Measuring Behavior Traits by Means of Repeated Short Samples," *Journal of Juvenile Research*, 12: 230-235 (1928).
4. HEINLEIN, C. P., "New Method of Studying the Rhythmic Responses of Children together with an Evaluation of the Method of Simple Observation," *Pedagogical Seminary and Journal of Genetic Psychology*, 36:205-228 (June, 1929).
5. MCGROCH, J. A., and WHITELY, P. L., "The Recall of Observed Material," *Journal of Educational Psychology*, 17: 419-425 (1926).
6. MÜNSTERBERG, H., *On the Witness Stand* (Doubleday, Page & Co., 1908; now published by Doubleday, Doran & Company).
7. NEWCOMB, T. M., *The Consistency of Certain Extrovert-Introvert Behavior Patterns in 51 Problem Boys*, Teachers College Contributions to Education, No. 382 (1929).
8. OLSON, W. C., *Measurement of Nervous Habits in Normal Children*, Institute of Child Welfare Monograph Series, No. 3 (University of Minnesota Press, 1929).
9. ———, "Method for Observational Measurement," *National Conference of Supervisors and Directors of Instruction, Scientific Method in Supervision, Second Yearbook* (1929), pp. 195-205.
10. SYMONDS, P. M., "Study Habits of High School Pupils—as Shown by Close Observation of Contrasted Groups," *Teachers College Record*, 27: 713-724 (April, 1926).
11. THOMAS, D. S., and others, *Some New Techniques for Studying Social Behavior*, Child Development Monographs, No. 1 (Teachers College Bureau of Publications, 1929).
12. WINCH, W. H., "Can Observation Be Trained in School Children?" *Journal of Educational Research*, 15:229-238, 314-326 (1927).

Chapter III

RATING METHODS

Purposes of Rating

RATING methods, which are among the recognized means of measuring conduct, are used both for purposes of investigation and in practical ways to evaluate personnel in industry and schools. Introduced at first timidly and with reservation, rating methods are now employed with growing confidence in large industrial, mercantile, and financial establishments. Reports have been made of the results of rating methods to assist in recommending promotions, determining wage increases, and deciding transfers to positions of greater authority. In education, ratings are used widely in school systems to evaluate personnel. Teachers' agencies quite generally resort to them in an endeavor to obtain information concerning their clients. Rating methods help a large organization keep the same contact with its personnel that may be obtained through personal observation in a small group. Systematic rating has been credited with the following advantages.

Advantages of Rating

Rating is an aid in administration. By means of a permanent and objective record, an administration is enabled to evaluate the worth of its personnel. A school superintendent or business manager, through the use of ratings, may be enabled to make wise decisions regarding employment, placement, transfers, promotions, and dismissals. Although ratings are imperfect, as will be subsequently shown, nevertheless by the use of various devices an administrator is able to make less subjective decisions than customarily result when the only evidence is that of casual observation, gossip, or the influence of personal favorites. Rating methods are not merely another device of the efficiency expert. They are

a means of getting important information that often can be gained in no other way.

Ratings stimulate the person being rated. Ratings stimulate the person being rated. It is often recommended that persons be shown the results of their rating. If this is done in a dignified, impersonal way, and if the person rated is made to feel that the ratings have been honestly given, and particularly that they represent the pooled judgment of several of his superiors, the result should be salutary. An employee is thereby made aware of his deficiencies and shortcomings and has every incentive to improve himself in those particulars wherein he falls short. Furthermore, if the employee has previously rated himself on the same scale and then compares his own estimate of himself with the estimate made by others, he can often be made to see clearly in what respects he has misjudged himself in comparison with the impression he makes on others. A sympathetic supervisor can often make ratings an effective tool in his delicate task of improving the efficiency and morale of personnel.

It should be added, however, that in certain successful personnel work in industrial and mercantile establishments rating methods have been found to be relatively inflexible and rigid in comparison with interview methods, especially when the latter are conducted from the psychiatric point of view.*

Ratings react in a favorable way on the person doing the rating. Ratings react in a favorable way on the person doing the rating. In drawing up the rating schedule he is made to think out closely just what qualities are desirable in his employees. He becomes more observant both of the desirable qualities and of the differences among his staff in respect to these qualities. He is made more sensitive to undesirable points in a man, but he is also quicker to recognize and reward praiseworthy traits. Because his attention is directed to the characteristics of his workers he becomes more sensitive in appreciating them, and the resulting in increased interest has its rebound in a better morale among the force.

Ratings, if periodically given, help keep alive the personnel spirit. They make a staff feel that they are being held responsible for good service. There is nothing quite so destructive as neglect. An unoccupied house apparently deteriorates much faster than a

* See Anderson, V. V., *Psychiatry in Industry* (Harper & Brothers, 1929).

house wherein people live. Failure to pay attention to personnel is sure to lead to carelessness, slovenliness, and loss of efficiency. In a large organization where the personal touch of the management can only extend a little way, ratings repeated periodically and systematically help to keep the personnel spirit alive. This is the more true if the force is given evidence that the ratings are not merely filed away, but are studied and used as a basis for decisions concerning the individuals rated.

Ratings help make judgment analytical. Without rating methods one is inclined to give an opinion concerning a person which is no more than a general impression. A rating schedule usually asks for opinions on several different traits or qualities. If thoughtfully selected, these separate qualities may be made to represent important sides of a man's personality. In using such a scale the rater is forced to consider one quality or trait at a time to the exclusion of others—he is forced to be analytical. At the end of the rating he has not expressed a single blanket opinion concerning a man, but has analyzed his opinion into many phases. No man is wholly or evenly good or bad in all respects, and the carefully planned rating schedule yields a judgment that indicates points of relative strength or weakness. A rating scale makes it possible to give analytical judgment.

Ratings systematically given tend to make judgment representative. If an opinion or a judgment of a man is given in an emergency or when the man is under special consideration, the rating is especially likely to be colored by the rater's likes and dislikes and his own interest in the situation. If an opinion is given after a man has made himself prominent by some word or act, the rating is sure to be warped. It will be a near-sighted estimate. Ratings should be made on no special occasion and under no unusual pressure, but regularly and systematically as a routine measure.

Finally, ratings are a recognized method of getting data for research purposes. In psychological research rating has already had wide use and has performed valuable service. Cleeton and Knight (21), for instance, resorted to ratings to demonstrate that character could not be read by external signs. In the field of social psychology and in the study of the personality there is often no other method of measurement available than rating. Ratings are used to validate more objective testing methods. The

rating data that accumulate in offices, industrial plants, and school systems may yield valuable information in research as to the importance of certain personal qualities in work of various kinds. Any attempt to develop objective testing methods for the measurement of personnel has to depend upon ratings for its validation.

Administration of Ratings

Ratings should be given periodically and systematically. Most rating systems to-day are carefully planned and provide that records be entered on a prepared chart. Very frequently a graphic chart (described later) is used. In introducing ratings on a large scale, plans should be laid for a uniform and standardized system. Simply to ask supervisors and managers for an offhand expression of opinion is certain to result in chaos, and the information they will turn in will have all of the defects of the most subjective evidence. In order to make ratings as accurate, objective, and free from bias as possible, the conditions under which they are made should be thoroughly standardized. Ratings should be made on cards of uniform size, for convenient sorting and filing. Uniformly printed rating blanks will make possible a comparison between different ratings and so enable the statistician to compile, summarize, and interpret the rating results of various groups at different times.

Uniform and standardized blanks should be used in rating. If uniform and standardized blanks are prepared, and if some graphic scheme of marking is adopted, the rating is easy. These are the important factors to be considered. Every rating represents an act of judgment, weighing, and comparison. Usually rating requires the making of fine distinctions on slender evidence; hence any device that can reduce the difficulty of the task is of value. Executives find the modern standardized graphic rating blanks easy to use. Raters, freed from thinking up things to say and searching for descriptive adjectives, can concentrate directly on the rating.

The best results are obtained if ratings are made periodically, either annually, semi-annually, or quarterly. They cannot be made too often, at least so it seems if we overlook the practical objections that time is required to make the judgments and that most people feel a natural reluctance to making judgments of this kind.

Of course in administering the system a definite date should be set on which the ratings are due.

Ratings should be made to yield quantitative scores. Although the use of descriptive adjectives such as *excellent*, *good*, *average*, *fair*, or *poor* smacks of subjective evaluation, it is probably true that a rating in terms of a number scale is no more objective than one in terms of descriptive adjectives. Indeed, the graphic scale does away with the use of numbers and permits the rater to think entirely in terms of descriptive adjectives, a far more familiar form of thought. But whether the rating is given in numerical terms or whether descriptive adjectives are translated into numbers, numbers are necessary if the ratings are to be given statistical treatment. Although for the more immediate uses of rating in personnel work statistical treatment is not necessary, whenever ratings are gathered with a possibility that they may be used in a research way they should be quantitative. A numerical rating scale makes it possible to check up on the accuracy of the raters.

Ratings, for use either in personnel work or in research, should be in the form of an objective record. All scientific work demands data that may be studied and worked over. In the natural sciences one may study actual materials or organisms. The chemist works with elements and compounds, and the biologist studies tissues under the microscope. And the psychologist, whose material is the transitory reaction, seeks for data in no less objective form. He is fortunate when the reaction leaves behind it a permanent record, as when a subject takes a paper and pencil test. At other times he is forced to make special arrangements to obtain a permanent record of the reaction made. A rating is a permanent, objective record of one's observation of conduct. The value of a rating is diminished only by the subjectivity of the observation which it records. In its use in personnel administration, an objective record of the rating has other values, such as the possibility of comparison with other records of the same individual obtained at other times and under other conditions; or by getting cumulative ratings, successive ratings may be compared with each other, growth may be noted, and progress may be determined. If ratings are worth making at all, they are worth keeping.

Besides the original rating sheets, which should always be filed away and kept, there may be several types of transcripts. In a

large organization a checking sheet should contain data as to the time when ratings are turned in, objective measures of the accuracy of the rating and notes as to the care and value of the ratings of each of the raters. Another sheet may contain a summary of the ratings of each employee so arranged as to show the comparison of ratings on any individual and the growth or change over a period of time. Ratings also should be entered on each employee's personnel record card, and here again cumulative records may indicate progress and enhanced value of the man to the company.

Observation and note-taking preliminary to rating. The best results in rating come from taking pains. One way to get poor ratings is to ask a person to sit down and give in an offhand, opportunistic way his opinion. Ratings, in order to be of value, must be made with more than ordinary care. In the first place the rater must be brought wholly into sympathy with his task. His attitude must not be that of passive acquiescence in the task of giving ratings; rather, he must take an active interest in assuring himself that his judgment is worth seeking, is merited.

Good rating demands a preliminary period of observation. During this observation the rater should be instructed to watch for certain definite things. Usually in rating persons the observation should be directed to a man's acts and his conduct. Even if the rating is to be on a man's qualities or traits, those qualities or traits should be analyzed into their behavior components, and these acts of conduct should be observed to aid in forming a judgment. Webb (120) recommends that the observation period should be extended not merely over hours or days, but over weeks and months. An hour's observation is certain to result in a narrow sampling of a man's conduct, and such ratings as are based on limited observations are sure to be warped. Webb also notes that the observations of different raters must be independent. Raters should not compare notes before giving their ratings, but wait until after their judgments are recorded. If the observations are not independent, the separate judgments will not count as single judgments, for one judgment will influence another. Notes, which are descriptions of actual conduct, should be taken during the period of observation, and as far as possible the rating should be based on the objective records which the notes contain. Good notes contain a minimum of generalization, inference, or opinion.

Before giving the rating the rater should dispassionately review the evidence contained in his notes based on observation. He should eliminate as far as possible personal bias, his likes and dislikes, common gossip, and reputation. The ideal of rating is a pure act of judgment based on whatever evidence he is able to gather. Thus by making long, careful observation, recording these observations when made, and basing judgment on what is observed, ratings are brought to their highest reliability. Quite often one is asked to rate on qualities or traits which one has never even had the opportunity to observe. Again one may be asked to rate on qualities or traits to which he has never given his attention. Only too frequently we are asked for facts for which the evidence has never been collected.

The training of raters. This is an obvious sequel to the need for observation in order to obtain good ratings. It is commonly yet erroneously assumed that any one can give ratings. The best results come after the raters have been given instructions in the art of rating and, as preliminary to this, in the art of observing and note-taking. They also need instruction on how to use the scale, how the scale is marked, how to give ratings that yield an unskewed distribution, and how to avoid the pitfall of allowing the general impression of a person to influence the rating on a specific trait. Raters need practice in rating and criticism of their ratings by a competent person.

Kingsbury (61) reports the results of experience in the training of raters in a large bank. He proposes the following procedure. First, there should be instruction in the purposes of rating and in the use of the particular rating scale adopted. Second, the rater should make a set of ratings. Third, these ratings should be analyzed by the psychologist. Fourth, the psychologist and rater should confer. Fifth, there should be reinstruction in which the particular faults and weaknesses of the rating are pointed out. Sixth, there should be rerating. Kingsbury states that this procedure was repeated three times before the raters were given a certificate of competence. Only after similar preparation can satisfactory ratings be obtained.

This particular institution prepared a rater's manual which was used in giving the instruction in rating. This manual contained the following topics.

1. Statement of the purposes of rating.
2. Exposition of the responsibility of raters.
3. Detailed description of rating scales.
4. Description of the procedure to be followed in making ratings.
5. Enumeration of common errors to be guarded against in rating.
6. Explanation of some of the methods used in analyzing ratings which it is suggested the manager may use in checking his own ratings.

The two main checks that can be readily and objectively obtained on a rating are *average error* and *direction of error*.

TABLE 3
COMPUTATION FOR DETERMINING AVERAGE ERROR AND DIRECTION OF
ERROR IN RATING

<i>Individuals rated</i>	<i>Average rating</i>	<i>Rating being studied</i>	<i>Differences without signs</i>	<i>Positive differences</i>	<i>Negative differences</i>
A	6.5	6.0	0.5		-0.5
B	2.6	4.0	1.4	1.4	
C	7.9	7.0	0.9		-0.9
D	3.2	3.0	0.2		-0.2
E	0.8	3.0	2.2	2.2	
F	4.5	5.0	0.5	0.5	
G	6.7	4.0	2.7		-2.7
H	8.4	9.0	0.6	0.6	
I	8.0	8.0	0.0		
J	2.2	3.0	0.8	0.8	
			9.8	5.5	-4.3
			$9.8 \div 10 = .98$		

Average error in rating..... 0.98

Rating too high by an average amount of..... 0.12

* The average rating is the average of the several supervisor's ratings on the individuals being rated.

The steps in the procedure are as follows:

1. Place the ratings of the supervisor who is being studied opposite the "true ratings," i.e., the average rating of all the supervisors as in the above example.
2. Take the difference without regard to sign and place in the fourth column.
3. Find the sum of this column.
4. Divide this sum by the number of individuals rated. This is the average error in rating.

5. If the rating is higher than the true (average) rating, place the difference in the column headed *Positive differences*. Place a plus sign before each of these differences. If the rating is less than the true (average) rating, place the difference in the column headed *Negative differences*. Place a minus sign before each of these differences.

6. Add each of these columns: Call the sum of the *Positive differences* column plus, and the sum of the *Negative differences* column minus.

7. Subtract the sum of these two columns. Give the difference the sign of the larger number.

8. Divide by the number of ratings made. If the rating is the same as the true rating, the difference is zero and will not belong in either the positive or the negative column. One divides by the total number of ratings in any case. If the result is plus, the rating has been too high on the average; if the result is minus, the rating has been too low.

Schools are cautiously trying out ratings for the purpose of getting a supplementary mark in the shape of a measure of citizenship or other character traits. Scattered bits of evidence would indicate that there are immense possibilities of value in this rating, although—except in one or two instances—this rating has hitherto never been undertaken seriously. Where ratings have been discredited, the reason is partly that they have never been given a thorough try-out. It has been taken for granted that any one can make a character rating. Yet rating the quality of an English composition, or of a specimen of handwriting, is a skilled art which is mastered only after considerable practice. Schools should not undertake more extensive ratings until they have pondered over the nature of the problem they are up against. Standardized blanks or record cards should be drawn up and printed on which to record teacher ratings of pupils, these ratings to be called for periodically. Teachers must be encouraged to make systematic observation and must be requested to make notes of the results of their observation. Ratings which have been made should be studied statistically to determine average error and direction of error of the ratings. Then teachers should be called into conference and given definite instructions as to how their ratings may be improved. An upper limit of error may be established and this diagnosis and instruction of the teachers

as raters should continue until all attain a proficiency within the limit of error allowed. If ratings are worth while at all, they are worth doing well.

Traits as a Scale of Measurement

Because we usually judge a person in qualitative terms, it comes as a distinct surprise to some of us to learn that it is possible to rate another person quantitatively. In the old days (which have not altogether gone!), the value of a man's services was reported by the use of descriptive adjectives. Some of the qualities ascribed to Colonel Lindbergh in promotion reports in the army as cited by the President were "purposeful," "serious," "deliberate," "stable," "efficient," "industrious." But even so remarkable a character as Lindbergh probably does not always show deliberation or industry to the same degree at all times or in all situations. Descriptive language rests almost wholly on an all-or-none basis, while conduct itself is not dichotomous (serious-frivolous, good-bad, etc.). Characterization is rough and crude when it is accomplished by means of descriptive adjectives.

In assigning ratings one comes to think of a trait as possessing degrees. It is as though a scale were constructed different points along which would represent different degrees of the trait. Even though there is no objective means for defining points on the scale, and descriptive adjectives must still be resorted to, the gain comes in recognizing that the trait may be present or absent in varying degrees. Suppose that the rater is giving his judgment as to the honesty of an individual. He does not pronounce him merely honest or dishonest, but while thinking of a scale running from the utmost dishonesty through varying degrees of less than perfect rectitude to perfect uprightness, he tries to place the individual somewhere along that scale, according to his judgment of the facts. In making such a judgment a rater tries to think of the *average* status of the person being rated in the trait under consideration, having in mind *no special instance* when the individual exhibited the trait in question.

It will be found that most personnel ratings are on traits rather than conduct habits, probably because traits are more comprehensive and general. We feel in rating traits that we are rating more fundamental and important characteristics of a person than

would be the case in rating merely on behavior or conduct. However, the reality of what is designated by the name of a trait has often been questioned. Is an individual always exactly honest in entering his golf score or in stating his son's age in buying a railroad ticket? If one stops to think just how the rating on some trait is made, only vague recollections are brought to mind. Wells (121, p. 15) tells us that judgments for which we can give no reason are often as good as those for which we can give a reason. Landis (72) has asked raters on what they based their judgments, with the interesting results that half of the reasons given for judgments were vague and non-specific, evidently made on vague, fleeting impressions—probably the residue of forgotten experiences. Landis found that when forced to justify a rating, raters give the most bizarre reasons. He also found that specific reasons were most often given when the rating implied some distinctly undesirable or unsocial trait. In other words the typical rating is not the result of a compilation of a series of clear, definite observations, but is a general impression based on a variety of half-forgotten contacts and experiences. Scanty evidence at best is all that one can recall when asked concerning a person's honesty: perhaps it is an instance of laxity on the part of a storekeeper in returning some overcharge that is remembered, or perhaps the misrepresentation of an event in a friend's account of it. In short, unless certain events stand out clearly in memory, one tends to class all people as about even in rectitude and trustworthiness. To give an adequate rating one should have noticed the individual carefully in many situations where the trait might be exhibited. For instance, if it is a question of honesty, one might observe and give answers to such questions as:

1. Does he return borrowed money promptly?
2. Does he return money when he receives too much in change?
3. In traveling does he pay his fare without any attempt to dodge?
4. Does he return found articles?
5. Does he keep a promise?
6. Does he obey rules?
7. Does he play games honestly?
8. Does he refrain from cheating in an examination?
9. Does he recite without attempting to bluff?
10. Does he do his work independently?

In short, if what is called conduct rating, as opposed to trait rating, is attempted, instead of judging traits one may rate the expected response in a particular situation. The answers to the above questions might be simply *yes* or *no*. But even a single and limited act of conduct may receive a more detailed rating, perhaps in terms of the different probabilities of such conduct, using adjectives such as *never*, *seldom*, *sometimes*, *often*, *always*. There is always a probability, ranging from zero to one, that a certain act will take place. Sometimes the act will take place only once out of 100 possible times, sometimes ten out of 100, sometimes ninety out of 100; and these chances can be estimated in rating.

There is undoubtedly a relationship between the rating of traits and the rating of conduct. The element in behavior is a reaction in a certain situation, and our primary judgments are always based on our observations of actual conduct. Just how far one's judgment of a trait is based on the observation and judgment of conduct it is impossible to state. Further, it remains for experimentation to prove that the average of the judgments on the above ten questions on honesty will correlate perfectly with the rating of the individual on honesty in general. Although close, they are probably never identical. Of the two, we should prefer as a measure of honesty the summation of answers to the ten questions, rather than a rating on honesty itself. For we do know that the answers to the ten questions represent a fair sampling of the field of honesty, whereas a blanket rating may rest on evidence derived from two or three scattered and partial observations.

General Methods of Making Ratings

Rating methods are almost as old as any experimental methods in psychology. The beginnings of rating methods can be traced back to the work of Fechner and Mantegazza in psychophysics. The first rating scale, in its modern sense, was that published by Galton in 1883 in a section on "Mental Imagery" in his *Inquiry into Human Faculty and Its Development** (pp. 64, 65). This scale even to-day may be considered a model of its kind.

* E. P. Dutton & Co., Inc.

Scale of Mental Imagery—Imagery of a Breakfast Table

- "Highest.—Brilliant, distinct, never blocky.
- "First Suboctile.—The image once seen is perfectly clear and bright.
- "First Octile.—I can see my breakfast table or any equally familiar thing with my mind's eye quite as well in all particulars as I can if the reality is before me.
- "First Quartile.—Fairly clear; illumination of actual scene is fairly represented. Well defined. Parts do not obstruct themselves, but attention has to be directed to different points in succession to call up the whole.
- "Middlemost.—Fairly clear. Brightness probably at least from one half to two thirds of the original. Definition varies much, one or two objects being much more distinct than the others, but the latter come out clearly if attention be paid to them.
- "Last Quartile.—Dim, not comparable to the actual scene. I have to think separately of the several things on the table to bring them clearly before the mind's eye, and when I think of some things the others fade away in confusion.
- "Last Octile.—Dim and not comparable in brightness to the real scene. Badly defined, with blotches of light; very incomplete; very little of one object is seen at one time.
- "Last Suboctile.—I am very rarely able to recall any object whatever with any sort of distinctness. Very occasionally an object or image will recall itself, but even then it is more like a generalized image than an individual one. I seem to be almost destitute of visualizing powers as under control.
- "Lowest.—My powers are zero. To my consciousness there is almost no association of memory with objective visual impressions. I recollect the table but do not see it."

Another famous scale is that set up by Pearson (85) for judging intelligence.

Pearson's Scale of Ability

- A. *Mentally defective.* Capable of holding in the mind only the simplest facts, and incapable of perceiving or reasoning about relationships between facts.
- B. *Slow dull.* Capable of perceiving relationship between facts in some few fields with long and continuous effort; but not generally nor without much assistance.
- C. *Slow.* Very slow in thought generally, but with time understanding is reached.
- D. *Slow intelligent.* Slow generally, although possibly more rapid in certain fields; quite sure of knowledge when once acquired.

54 Diagnosing Personality and Conduct

- E. *Fairly intelligent.* Ready to grasp, and capable of perceiving facts in most fields; capable of understanding without much effort.
- F. *Distinctly capable.* A mind quick in perception and in reasoning tightly about the perceived.
- G. *Very able.* Quite exceptionally able intellectually, as evidenced either by the person's career or by consensus of opinion of acquaintances, or by school record in case of children.

In using a scale of this kind one uses a simple ruled sheet on which are listed the names of those for whom ratings are desired. If the different levels of the scale are numbered the appropriate numbers may be entered against each name to indicate the rater's judgment.

RATINGS ON INTELLIGENCE		Rating
Smith	3
Jones	3
Brown	4
Thompson	1
Black	6

Other devices for recording such ratings will suggest themselves to the ingenious psychologist. It is always possible to let the process of rating be that of checking or underlining or crossing out instead of writing in numbers, as shown in the illustration.

RATINGS ON INTELLIGENCE									
Smith	1	2	③	4	5	6	7	
Jones	1	2	③	4	5	6	7	
Brown	1	2	3	④	5	6	7	
Thompson	①	2	3	4	5	6	7	
Black	1	2	3	4	5	⑥	7	

As against these methods of rating, probably the simplest methods which could be devised, it has been suggested that in rating men on any given trait much longer descriptions of different positions on the scale might be advantageously used. For example, each scale point could be illustrated by describing in more or less detail and in a vivid way actual individuals. The suggestion is prompted by the hypothesis that by thus more clearly defining each point on the scale the ratings could be given more objectively and hence more reliably.

Another method of giving ratings is called the *method of paired comparisons*. In this method the rater compares each individual being rated with every other individual being rated with respect

to the trait under consideration. The comparison is simply in terms of better or worse. A good scheme for using this method is to draw up a ruled table upon which each comparison is indicated by placing a cross in the appropriate square to designate superiority. First compare individual A with the other individuals, placing a cross in the A row under B if A is judged superior to B, etc. The marked chart indicates how this scheme is carried out.

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>Totals</i>
<i>A</i>		X	X		X	3
<i>B</i>			X		X	2
<i>C</i>					X	1
<i>D</i>	X	X	X		X	4
<i>E</i>						0

The table should be made consistent by seeing that there are no conflicting judgments. If in the upper right section A is judged superior to E, then E must not be judged superior to A in the lower left section. Finally the number of crosses in each row may be summed across for a composite valuation.

Another method in common use is called the *order-of-merit* method, or more simply *ranking*. In this method the persons being rated are assigned serial numbers from 1 up. The most superior individual is given 1, the next individual is given 2, etc. If two individuals are judged to be tied for a given rank, they are both given a rank which strikes an average of their rank and the next rank following. For instance, if two individuals tie for first place, both are assigned the rank of 1.5, and the next individual is ranked 3. If three individuals tie for first place, all three are given the rank of 2, and the next individual is given the rank of 4. Ranking is expedited by writing the names to be ranked on cards, which may be rearranged in the order desired.

Still another method of rating is the *score card*. Whatever is being rated, whether it be the quality of school buildings, or the efficiency of teachers, or the merit of textbooks, is analyzed into constituent parts to each of which a maximum score is assigned. These allotments of maximum score are usually determined by asking various competent persons to analyze whatever is being

rated and to assign relative weights according to their judgment. These weights are then averaged. Usually the score card is so constructed that the possible total of maximum credits is 100 or 1,000. In using the score card the rater who is judging a specific textbook or school building or teacher assigns a value to each item which is some proportional part of the maximum credit allowed that item in the score card. These credits are then totaled for the final score. The difficulty in this method, which is also a limitation upon the adequacy of the method, is in assigning the credits to the separate items.

The score card given here for illustration aims to measure the merit of church school textbooks. It was constructed by C. C. Peters and is to be found in Volume II of the *Indiana Survey of Religious Education* (pp. 109-114), Institute of Social and Religious Research, W. S. Athearn, editor.

SCORE CARD FOR MEASURING THE MERIT OF
CHURCH SCHOOL TEXTBOOKS

	POINTS Main Headings	POINTS Sub-Headings
I. Mechanical Features	115	
1. Type		25
2. Attractiveness of page		20
3. Pictorial illustrations		28
4. Organization of page		21
5. Make-up of book or pamphlet		20
II. Style	100	
1. General literary merit		45
2. Appropriateness of style to age of pupils		55
III. Pedagogical Organization of Lessons	250	
1. Organization of the lesson about an aim		56
2. Type of organization of the lesson		41
3. Provision for controlling study		50
4. Provision of means to insure functioning of the instruction		65
5. Provision for the enrichment of experience in ways not directly related to the lesson aim but not antagonistic to it		38
IV. Teaching-Helps in the Individual Lesson	140	
1. A separate manual for teachers		32
2. Valuable supplementary material for teachers		31
3. Useful teaching suggestions		38
4. Valuable teaching aids		39

	POINTS	POINTS
	<i>Main Headings</i>	<i>Sub-Headings</i>
V. Teachings-Helps Involved in the Organization of the Book as a Whole	125	
1. Valuable teaching suggestions additional to those that constitute an integral part of each lesson (as in introductory chapter or scattered in short notices through the book)		34
2. Supplementary teaching material		38
3. Provisions for giving the teacher perspective of the course		29
4. Provision for review lessons		24
VI. Content	270	
1. Fitness of the material to appeal strongly to pupils of the age for which the lesson is intended		95
2. Fitness of the material to meet the needs of the pupils as defined by child psychology and by sociology (age-levels considered)		110
3. Fitness to meet the specific objectives of the particular church (or other group) for which the material has been prepared		65

Man-to-Man Rating

One of the significant developments of the rating scale technique during the World War was the man-to-man scale. The Army Rating Scale, reproduced below, made use of this principle. It has often been urged that in rating one should not compare a man with abstract qualities defined on paper, but should make a direct comparison of one man with another. Rating should not consist in the assignment of a man to one out of a set of compartments labeled *excellent*, *good*, *average*, *poor*, or *bad*, but should be a direct comparison of one person with another. The question asked should be, "Is Jones a better man than Smith?" in the trait being considered.

The man-to-man rating principle was developed by Dr. Walter Dill Scott and his associates in the Bureau of Salesmanship Research of the Carnegie Institute of Technology. The Army Rating Scale was first framed by Scott in May, 1917. The Army Rating Scale which is reproduced below contains the five headings Physical Qualities, Intelligence, Leadership, Personal Qualities, and General Value to the Service. For each quality an officer was to

select five men of his acquaintance, one possessing the quality in its greatest possible degree, one midway between this maximum and the average, one average in the quality, one midway below the average and the lowest degree and one possessing a minimum of the quality. These "scale men" constituted the scale by means of which all other men were rated by a direct comparison.

How to Make the Scale (92, pp. 203-206)*

"1. Write on small slips of paper the names of from 12 to 25 officers of your own rank and not above the average age of that rank. They should be men with whom you have served or with whom you are well acquainted. Include officers whose qualifications are extremely poor as well as those who are highly efficient. If these names do not include all the grades, for each of the five qualifications, others may be added.

"2. Look over your names from the viewpoint of Physical Qualities only. Disregard every other characteristic of each officer except the way in which he impresses his men by his physique, neatness, voice, energy, and endurance. Arrange the names on the slips of paper in order from highest to lowest on the basis of the physical qualities of the men. Select that officer who surpasses all the others in this qualification and enter his name on the line marked Highest under Physical Qualities. Then select the one who most conspicuously lacks the qualities and enter his name on the line marked Lowest. Select the officer who seems about half way between the two previously selected and who represents about the general average in physical qualities; enter his name on the line marked Middle. Select the officer who is half way between Middle and Highest; enter his name on the line marked High. Select the one who ranks half way between Middle and Lowest; enter his name on the line marked Low.

"3. In the same manner make out scales for each of the other four qualifications (Intelligence, Leadership, Personal Qualities, and General Value to the Service).

"4. Each officer whose name appears on the Scale should be one who exhibits clearly and distinctly the qualification and the degree of the qualification for which he has been chosen.

"5. The names for Highest and Lowest on each section of the Scale must represent extreme cases. The name for the Middle should be that of an average officer, half way between extremes. High and Low should be half way between the Middle and the extremes. An even gradation of merit is important.

* Scott, W. D., and Clothier, R. C., *Personnel Management* (A. W. Shaw Company, 1923), pp. 203-206. By permission of the present publishers, McGraw-Hill Book Company.

"6. In making or using any section of the Scale, consider only the qualification it covers, totally disregarding all the others.

"7. In rating subordinates of more than one grade, the best practice is to make separate scales for each grade, using always the names of officers one grade higher than that of the subordinate to be rated. However, in exceptional cases good results have been secured where a scale constructed of captains is used for rating both lieutenants and captains, and a Scale constructed of colonels is used for rating all ranks of field officers.

How to Use the Scale

"8. Rate your subordinate for Physical Qualities first. Consider how he impresses his men by his physique, bearing, neatness, voice, energy, and endurance. Compare him with each of the five officers in Section I of your Rating Scale, and give him the number of points following the name of the officer he most nearly equals. If he falls between two officers in the Scale, give him a number accordingly (e.g., if between Low and Middle, give him, 7, $7\frac{1}{2}$, or 8).

"9. Rate the subordinate in a corresponding manner for each of the other four essential qualifications. Under III (Leadership) and V (General Value to the Service), consider which officer he will most nearly equal after equivalent experience.

"10. In rating, make a man-to-man comparison of the subordinate with the officers whose names appear on your Scale—never in terms of numbers directly. Disregard the numerical equivalent until you have made these concrete comparisons.

"11. When rating several subordinates, rate all of them on each qualification before adding the total for any one.

"12. This is not a percentage system and you should not allow yourself to fix in mind any particular number of points you think the subordinate ought to get.

"13. The total rating for a subordinate is the sum of the ratings you give him in the five separate qualities. If directions are followed carefully, the average of any considerable group of officers rated is about 60 points. In other words, 60 points for a lieutenant means that a captain has compared him with the captains he knows and certified that after equivalent experience he will be equal to an average captain.

"14. Each officer below the rank of Brigadier General will be rated by his immediate superior. Ratings will be revised or approved by the immediate superior of the officer making the rating. The revising officer will use his own scale and make ratings independently of those made by the rating officer. Superior officers will see that their subordinates make all ratings according to the Rating Scale system, in order that a just and equitable record may be had for all officers in the Army.

Diagnosing Personality and Conduct

Army Rating Scale

"I. PHYSICAL QUALITIES

"Physique, bearing, neatness, voice, energy, and endurance.
(Consider how he impresses his men in the above respects.)

Highest	Captain.. Benson.....	15
High	Lieutenant.. Gray.....	12
Middle	Lieutenant.. Spencer.....	9
Low	Captain.. Clark.....	6
Lowest	Lieutenant.. Anderson.....	3

"II. INTELLIGENCE

"Accuracy, ease in learning, ability to grasp quickly the point of view of commanding officer, to issue clear and intelligent orders, to estimate a new situation, and to arrive at a sensible decision in a crisis.

Highest	Captain.. Clark.....	15
High	Captain.. Benson.....	12
Middle	Lieutenant.. Jones.....	9
Low	Captain.. Jorgensen.....	6
Lowest	Lieutenant.. Wilson.....	3

"III. LEADERSHIP

"Initiative, force, self-reliance, decisiveness, tact, ability to inspire men and to command their obedience, loyalty and coöperation.

Highest	Lieutenant.. Spencer.....	15
High	Captain.. Jorgensen.....	12
Middle	Captain.. Mathewson.....	9
Low	Captain.. Benson.....	6
Lowest	Lieutenant.. Alexander.....	3

"IV. PERSONAL QUALITIES

"Industry, dependability, loyalty, readiness to shoulder responsibility for his own acts, freedom from conceit and selfishness, readiness and ability to coöperate.

Highest	Lieutenant.. Wilson.....	15
High	Lieutenant.. Spencer.....	12
Middle	Lieutenant.. Alexander.....	9
Low	Lieutenant.. Anderson.....	6
Lowest	Lieutenant.. Jones.....	3

"V. GENERAL VALUE TO THE SERVICE

"His professional knowledge, skill and experience; success as an administrator and instructor; ability to get results."

Highest	<i>Captains.. Abbott.</i>	15
High	<i>Lieutenant.. Alexander.</i>	12
Middle	<i>Captains.. Mathewson.</i>	9
Low	<i>Captains.. Clark.</i>	6
Lowest	<i>Lieutenant.. Peters.</i>	3

The rating scale is a master scale like a yardstick or thermometer with which the men to be rated are directly compared.

One difficulty with the use of the scale is that each man's scale is his own and may have only slight relation to the scale of some one else rating the same individuals. For instance, the average man on A's scale may be distinctly superior to the average man on B's scale. Consequently the man-to-man method does not avoid errors due to tendencies to rate too high or too low. Paterson and Ruml (84) emphasize the necessity for calibration of the "scale men" when the man-to-man scale is used extensively. A similar trouble develops because scale men do not necessarily represent equal distances. Cleeton and Knight (22) make the point that man-to-man rating tends to avoid overestimation because one is making a direct comparison of individuals. Overestimation, however, can be surmounted by defining the per cent or number of individuals to fall in each class, or it may be corrected after the ratings have been made.

The man-to-man scale has not had extensive use since the World War. Concerning this Scott and Clothier (92, p. 207) say,

"The army form of Rating Scale has been found generally inapplicable to business and industrial conditions by virtue of the fact that it is relatively cumbersome in use. The construction of the master scale and the mental balancing of one man against another call for an expenditure of time and effort which the average executive is not in a position to contribute. It was found that in addition to fulfilling the need for accuracy, the industrial rating scale must fulfil the need for ease of operation. The success of any rating procedure necessarily depends upon the good-will and the intelligent coöperation of the executives and foremen under pressure of daily routine. It is very difficult to win this good-will and coöperation when the mechanical difficulties involved in the use of the scale are great." *

* Scott, W. D., and Clothier, R. C., *Personnel Management* (A. W. Shaw Company, 1923), pp. 207-208. By permission of the present publishers, McGraw-Hill Book Company.

The Graphic Rating Scale

The graphic rating method, to which reference has been made previously, is probably the most serviceable and widely adopted method of rating. The cross-on-a-line method has been in use for some time. Boyce (11) used this method prior to 1915, although his is not a true graphic rating plan, inasmuch as he divided his line up into intervals so that the rater merely crossed the line to indicate in which interval his judgment fell. The true graphic rating scale developed by the Scott Company Laboratory consists of a straight line about four or five inches in length beneath which are written descriptive adjectives or phrases to help define points along the scale. The line is intended to represent the scale of the quality or trait being rated, extending from one extreme of the quality to the other. The scale is marked by placing a cross on the line anywhere. It must be emphasized in giving the directions for using the graphic scale that the cross may be placed anywhere along the line, not necessarily above one of the descriptive adjectives defining the points along the scale.

The graphic rating scale may be scored or evaluated by means of a homemade rule or scale. A piece of cardboard may be used. Take a straight edge of the cardboard and mark off points along the edge to include a length equal to the length of the line of the rating scale. This space on the cardboard may then be divided up into intervals. As was argued on pages 78 and 79, seven intervals are sufficient, though any number of intervals may be used. To the writer's knowledge graphic rating scales have been constructed with thirteen, eighteen, twenty, and twenty-two intervals, and doubtless examples could be found with other numbers of intervals.

	1	2	3	4	5	6	7	8	9
Directions: Place the stencil so that the scale coincides with the graphic line. Note in which division the check falls. Enter this number in the column at the right of the line.									

The directions written on the stencil indicate how it is to be used. There may be an advantage in keeping the graphic principle of *crossing* a line, but the line may be pre-marked into seven

divisions so that the rater merely has to place his cross in one or another division.

Two graphic rating charts employing the graphic principle are reproduced here.

Rating Chart A

Pupil's Name Date.....
 School Grade.....
 Rated by

Directions for Using the Rating Chart

1. Let these ratings represent your own judgment. Do not confer with any one in making them.

2. In each trait or characteristic named below compare this pupil with the average pupil of the same age.

3. In rating for any particular trait disregard every other trait except that one. Many ratings are rendered valueless because the rater allows himself to be influenced by a generally favorable or unfavorable impression which he has formed of the person rated. Do not rate a pupil high on all traits simply because he is exceptional in some. Children are often very high in some traits and low in others.

4. Place a cross somewhere on the line running from "very high" to "very low" to indicate this child's standing in each quality. You may place your cross at *any point* on the line. It is not necessary to locate it at any of the division points or above any descriptive phrase.

5. Do not study too long over any one child. Give for each the best judgment you can, and go on to the next.

6. Give a rating for every trait.

7. The ratings will be held strictly confidential.

Health—Is he generally healthy and vigorous?

Bad	Poor	Average	Good	Excellent
-----	------	---------	------	-----------

Leadership—Does he take the lead in school affairs or does he follow others?

Always follows others	Rather tends to follow	Average	Rather tends to be a leader	Masterly, not easily influenced
--------------------------	---------------------------	---------	--------------------------------	---------------------------------------

The usual method with such a scale is to go through the traits for one individual and then turn the sheet to the next individual. Even though instructions are given to rate all individuals

on one trait before turning to the next trait, no direct comparisons can be made, since only one page is visible at a time. To make use of the close comparison which is demanded in ranking, rating scheme B is recommended:

Rating Chart B

Rated by Date.....
School Grade.....

Directions for Using the Rating Chart

1. Let these ratings represent your own judgment. Do not confer with any one in making them.
2. In each trait or characteristic named below compare this pupil with the average pupil of the same age.
3. In rating for any particular trait disregard every other trait except that one. Many ratings are rendered valueless because the rater allows himself to be influenced by a generally favorable or unfavorable impression which he has formed of the person rated. Do not rate a pupil high on all traits simply because he is exceptional in some. Children are often very high in some traits and low in others.
4. Place a cross in one of the compartments running from "very high" to "very low" to indicate each child's standing in this quality.
5. Do not study too long over any one child. Give for each child the best judgment you can, and go on to the next.
6. Give a rating for each child.
7. The ratings will be held strictly confidential.
8. Try to let the percentages guide you as to the number of crosses to fill in each compartment.

TRAIT: HEALTH

Is he generally healthy or vigorous?

Pupil	4% Very bad	11% Bad	21% Poor	28% Average	21% Good	11% Very good	4% Excel- lent
Charles							
William							
George							
etc.							

Scale B is generally to be preferred because it keeps the desirable features of the graphic rating chart and also permits the

close comparison of ranking. To be sure, the ratings for any one individual are spread over several sheets, but, as in any group test, the ratings for the several tests for any individual may easily be brought together when the numerical values of the graphic rating are measured.

Freyd (32, pp. 93, 94) lists the following advantages of the graphic rating method.

"It is simple and easily grasped.

"It is interesting and requires little motivation of the rater.

"It is quickly filled out.

"It frees the rater from direct quantitative terms.

"It enables the rater, nevertheless, to make his discrimination as fine as he cares, although this discrimination is lost if a scoring stencil of only a few points is used. [This advantage is somewhat illusory inasmuch as discrimination soon reaches a natural limit.]

"It is universal; that is, no master scale is required as in the Army Rating Scale.

"The fineness of the scoring method may be altered at will, yielding scores of from 1 to 5, or from 1 to 100.

"It allows of comparable ratings without requiring each rater to know all the members of the group."

Freyd (32, pp. 99, 100) gives a list of rules for making the graphic rating scale. Some of them, already mentioned in the earlier part of this chapter, will not be repeated. Rules which apply particularly to the graphic rating scale are:

"Decide on the extremes of the trait. It is frequently the case that one extreme of a scale may have several opposites.

"It will be found good practice to introduce every scale with a question, to which the rating furnishes the answer; for instance, the question 'How tactful is he?' or 'Is he tactful or tactless?' may be answered by checking on the rating line.

"The rating line should be of such a length that a stencil for scoring the rating can easily be calibrated.

"There should be no breaks or divisions in the line.

"The line should not be much more than five inches in length, so that it may be grasped as a unit.

"There should not be more than five descriptive items nor less than three.

"The end phrases of the scale should not be so extremely worded as never to be employed.

"The phrase descriptive of the neutral or average degree of the trait should be in the center of the scale.

"If there are five items, the intermediate ones should be closer in meaning to the center one than to the extremes. This has the effect of spreading the distribution. The same end may be accomplished by making the intervals on the scoring stencil smaller in the center than at the ends.

"Only universally understood phrases should be used. Slang is effective if there is no doubt as to its meaning.

"The descriptive phrases should be set in small type, and there should be plenty of white space between them.

"The favorable extremes of the scales should be alternated so as to do away with a tendency to check at one margin of the page."

The graphic rating scale which follows is taken from Filer and O'Rourke (31, p. 519):

Graphic Rating Scale

1. It is requested that you indicate by check (✓) your opinion of the applicant in each of the qualities specified. Place only one check after each quality. For example, on the specimen scale below, the check mark indicates that the supposed applicant is in the class which "Learns and adapts slowly" but is more nearly average than dull, because the check is placed nearer the "average" group than the "dull."

ABILITY TO LEARN. Consider ease and rapidity of under- standing new in- structions and adapting to new situations.	SPECIMEN—DO NOT MARK HERE				
	V				
	Dull and unadaptable.	Learns and adapts slowly.	Average in learning and adapt- ing.	Learns and adapts readily.	Learns with ex- ceptional ease and rapidity.

- Answer All of the Following

In giving your opinion on a particular trait, disregard for the moment every trait but that one, as specifically defined, and consider the applicant's ability in this trait from the point of view of GENERAL CLERICAL work only.

(a) ABILITY TO LEARN. Consider ease and rapidity of understanding new instructions and adapting new situations.	Dull and unadaptable.	Learns and adapts slowly.	Average learning and adapt- ing.	Learns and adapts readily.	Learns with ex- ceptional ease and rapidity.
---	--------------------------	---------------------------------	---	----------------------------------	--

(b) INDUSTRY. Consider energy and application to duties day in and day out.	Lazy.	Indifferent.	Average application.	Industrious.	Unusually energetic.
(c) INITIATIVE. Consider ability to go ahead with work without being told every detail, and to make practical suggestions for doing work in a better way.	Needs constant direction.	No originality.	Minor constructive ability.	Considerable constructive ability.	Highly constructive.
(d) COOPERATIVENESS. Consider ability to maintain good working relations with co-workers.	Troublemaker.	Causes slight friction.	Indifferent.	Cooperative.	Exceptionally cooperative.
(e) ATTITUDE TOWARD WORK. Consider voluntary interest and effort in work.	Unconcerned and no voluntary effort.	Interest and effort below average.	Average interest and effort.	Interest and effort above average.	Shows keen interest and wholehearted effort.
(f) SPEED. Consider rate at which applicant is able to work.	Very slow.	Slow.	Average rate.	Fast.	Exceptionally rapid.
(g) ACCURACY. Consider ability to do work without errors.	Unsatisfactory.	Makes many errors.	Average accuracy.	Seldom makes errors.	Exceptionally accurate.
(h) DISPOSITION. Consider natural temper of mind.	Decidedly ill-natured; uncivil.	Easily vexed; moody.	Average self-restraint.	Rarely vexed.	Exceptional self-control.

Diagnosing Personality and Conduct

(i) NEATNESS. Consider orderliness in work.	Disorderly	Somewhat below average in orderliness	Average orderliness	Somewhat above average orderliness	Exceptionally orderly.
(j) ABILITY TO SUPERVISE. Consider ability to direct work of others effectively.	Unable to direct work of others.	Somewhat below average ability in directing others.	Shows average ability in directing others.	Somewhat above average ability in directing others.	Maintains loyal and effective working force.

In their "Behavior Problem Record," the graphic rating scale has been adapted by Haggerty-Olson-Wickman* for use in recording judgments directly as to the frequency of certain types of behavior.

SCHEDULE A: BEHAVIOR PROBLEM RECORD**

DIRECTIONS FOR USING

Schedule A

Below is a list of behavior problems sometimes found in children. Put a cross (X) in the appropriate column after each item to designate how frequently such behavior has occurred in your experience with this child. A cross should appear in some column after each item. The numbers are to be disregarded in making your record. They are for use in scoring.

BEHAVIOR PROBLEMS	FREQUENCY OF OCCURRENCE				SCORE
	Has never occurred	Has occurred once or twice but no more	Occasional occurrence	Frequent occurrence	
Disinterest in school work	0	4	6	7	
Cheating	0	4	6	7	
Unnecessary tardiness ...	0	4	6	7	
Lying	0	4	6	7	
Defiance to discipline	0	4	6	7	
(Etc. for fifteen items)	0	4	6	7	

Total Score

* Olson, W. C., *Problem Tendencies in Children* (The University of Minnesota Press, 1930).

Haggerty, M. E., "The Incidence of Undesirable Behavior in Public School Children," *Journal of Educational Research*, 12:102-122 (Sept., 1925).

** From Haggerty, Olson, and Wickman's *Haggerty-Olson-Wickman Behavior Rating Schedules*. Copyright, 1930, by the American Council on Education and published by World Book Company, Yonkers-on-Hudson, New York.

SCHEDULE B: BEHAVIOR RATING SCALE*

DIRECTIONS FOR USING

Schedule B

1. Do not consult any one in making your judgments.
2. In rating a person on a particular trait, disregard every other trait but that one. Many ratings are rendered valueless because the rater allows himself to be influenced by a general favorable or unfavorable impression that he has formed of the person.
3. When you have satisfied yourself as to the standing of this person in the trait on which you are rating him, indicate your rating by placing a cross (X) immediately above the most appropriate descriptive phrase.
4. If you are rating a child, try to make your ratings by comparing him with children of his own age.
5. The masculine pronoun (he) has been used throughout for convenience. It applies whether the person whom you are rating is male or female.
6. In making your ratings, disregard the small numbers which appear below the descriptive phrases. They are for use in scoring.

DIVISION 1

1. How intelligent is he?

Score

People-minded (5)	Dull (4)	Equal of average child of street (3)	Bright (2)	Brilliant (1)
----------------------	-------------	---	---------------	------------------

2. Is he abstracted or wide awake?

Continually absorbed in himself (5)	Frequently becomes abstracted (4)	Usually present-minded (3)	Wide-awake (2)	Keenly alive and alert (1)
--	--------------------------------------	-------------------------------	-------------------	-------------------------------

3. Is his attention sustained?

Distracted: Jumps rapidly from one thing to another (5)	Difficult to keep at task until completed (4)	Attends adequately (3)	Is absorbed in what he does (2)	Able to hold attention for long periods (1)
--	--	---------------------------	------------------------------------	--

4. Is he slow or quick in thinking?

Extremely slow (5)	Sluggish, Floundering (4)	Thinks with ordinary speed (3)	Agile-minded (2)	Exceedingly rapid (1)
-----------------------	------------------------------	-----------------------------------	---------------------	--------------------------

5. Is he slovenly or careful in his thinking?

Very slovenly and illogical (5)	Inexact, A dabbler (4)	Moderately careful (3)	Consistent and logical (2)	Precise (1)
------------------------------------	---------------------------	---------------------------	-------------------------------	----------------

(Etc. for thirty-five items.)

Total, Division 1

* From Haggerty, Olson, and Wickman's *Haggerty-Olson-Wickman Behavior Rating Schedules*. Copyright 1930 by the American Council on Education and published by World Book Company, Yonkers-on-Hudson, New York.

PERSONALITY REPORT

Name of student.....

A. How are you and others affected by his appearance and manner?	Avoided by others	Tolerated by others	Liked by others	Well liked by others	Sought by others	No opportunity to observe

Please record here instances that support your judgment.

B. Does he need constant prodding or does he go ahead with his work without being told?	Needs much prodding in doing ordinary assignments	Needs occasional prodding	Does ordinary assignments of his own accord	Completes suggested supplementary work	Seeks and sets for himself additional tasks

Please record here instances that support your judgment.

C. Does he get others to do what he wishes?	Probably unable to lead his fellows	Lets others take lead	Sometimes leads in minor affairs	Sometimes leads in important affairs	Displays marked ability to lead his fellows; makes things go

Please record here instances that support your judgment.

D. How does he control his emotions?	Too easily moved to anger or fits of depression, etc.	Tends to be over emotional	Usually well balanced	Well balanced	Unusual balance of responsiveness and control
	Unresponsive, apathetic	Tends to be unresponsive			

Please record here instances that support your judgment.

E. Has he a program with definite purposes in terms of which he distributes his time and energy?	Aimless trifler	Aims just to "get by"	Has vaguely formed objectives	Directs energies effectively with fairly definite program	Engrossed in realizing well-formulated objectives

Please record here instances that support your judgment.

Another sample of the graphic rating scale is the "Personal-ity Rating Scale" * shown on pages 70 and 71, devised by a committee of the American Council on Education for use in colleges. This scale recognizes many of the safeguards and cautions which its makers had discovered in their experience with ratings, among them being: "Only traits observed by the rater should be rated; only traits for which no valid objective measurements are now available should be rated; the number of traits to be rated should not exceed five, if teachers are to be expected to rate the traits of a large number of students; traits should be mutually exclusive; a trait should not involve unrelated modes of behavior." Careful study of a large number of rating scales actually in use resulted in the selection of five qualities to be included in the scale which may be named as follows: Personal Charm, Initiative, Leadership, Emotional Control, Responsibility.

The Check List

This device, originally found promising by Hepner (44), was developed by Hartshorne and May (40) with considerable success. One hundred and sixty words descriptive of approval or disapproval were employed, and were divided roughly into two lists of eighty words each. For each word in one list its antonym appeared in the other list, but both positive and negative words appeared on both lists. Samples of the words are as follows:

Brutal	Undiscerning
Resourceful	Shirker
Courteous	Virtuous
Frivolous	Dignified
	etc.

A teacher in rating a pupil reads through the list and places a check mark against each word which can honestly be applied to the person in question. The directions require that the teacher consider each word, checking as many as she wishes to, or checking none at all if none apply. In scoring, the number of negative words checked is subtracted from the number of positive words checked, and the resulting score comes out either + 1, 0, or - 1 according as there is a balance of positive or negative words

* Robertson, D. A., chairman of Committee on Personality Measurements, "Personnel Methods," *The Educational Record*, Supplement 9: no. 8, July, 1928, p. 54.

checked. By a scoring method adopted later, the words in the lists are weighted empirically and more detailed scores are used. The reliability of the technique is reported by Hartshorne and May as .88 for total reputation, .74 when used to rate service, .64 when used to rate persistence, and .76 when used to rate inhibition.

The "Guess Who" Test

Hartshorne and May have devised a most promising technique, utilizing to good advantage a fact which is to be brought out later in the discussion of reliability, namely, that extreme ratings are especially reliable. The directions for using this device, called the "Guess Who Test," are as follows (40, p. 87):

"Here are some little word-pictures of children you may know. Read each statement carefully and see if you can guess who it is about. It might be about yourself. There may be more than one picture for the same person. Several boys or girls may fit one picture. Read each statement. Think over your classmates, and write after each statement the names of any boys or girls who may fit it. If the picture does not seem to fit any one in your class, put down no names, but go on to the next statement. Work carefully, and use your judgment.

"1. Here is the class athlete. He (or she) can play baseball, basketball, and tennis, can swim as well as any and is a good sport."

(2., etc.)

Several methods of scoring the device were tried, both by Hartshorne and May, and later by the writer. The method which seemed to yield the best results was the simple one of merely summing the total number of times a pupil is mentioned for any of the descriptions, giving "good" items a positive value and "bad" items a negative value. It might be expected that some one pupil in a class would be recognized by so many classmates as fitting a description that he would receive a preponderance of the votes, but actual experience shows that many children are usually recognized for a given description. It was also found that a child who received a large number of votes on any one question tended also to be mentioned for a large number of items. The results yielded a symmetrical distribution even though slightly peaked. Hartshorne and May report a reliability of .95 on the total test and .88

	HOW I FEEL	HOW MOST BOYS FEEL	HOW I THINK I OUGHT TO FEEL
Having other folks praise me	Dis- like Rather Don't Care Like Some a Lot	Dis- like Rather Don't Care Like Some a Lot	Dis- like Rather Don't Care Like Some a Lot
Reciting in class at school	Dis- like Rather Don't Care Like Some a Lot	Dis- like Rather Don't Care Like Some a Lot	Dis- like Rather Don't Care Like Some a Lot
Washing dishes	Dis- like Rather Don't Care Like Some a Lot	Dis- like Rather Don't Care Like Some a Lot	Dis- like Rather Don't Care Like Some a Lot

when measuring service. The writer found a reliability of .88 in his use of this device.

Self-Ordinary-Ideal Rating

The tendency to overrate oneself on desirable traits (discussed in greater detail on page 109) suggested to Knight and Franzen (66) a method of measuring various phases of personal maladjustment. G. B. Watson experimented further with the method, and finally Sweet (100) made a comprehensive analysis of its possibilities. His studies have resulted in the "Personal Attitudes Test." A better name to distinguish the method might be "Self-Ordinary-Ideal Rating," although in use with boys as subjects this name would lack the merit of disguising the nature of the test possessed by Sweet's title.

Many methods for recording the items could be devised. The first item is marked as one boy might have marked it. In Sweet's test the items were printed as they appear on this page (100, p. 6).

The test is scored in seven different ways to yield scores that are called "Self-Criticism," "Criticism of the Average Boy," "Feeling of Difference from the Average Boy," "Feeling of Superiority," "Feeling of Inferiority," "Deviation from Accepted Ideas of Right," and "Social Insight." In order to make the explanation of these scores as simple as possible, ratings on "How I Feel" will be called "Self"; ratings on "How Most Boys Feel" will be called "Ordinary"; and ratings on "How I Think I Ought to Feel," "Ideal."

Scores for *Self-Criticism* are found by counting the number of times "Ideal" is different from "Self." The higher the score, the more the boy tends to depreciate himself; the lower the score, the more perfect the boy says he is.

Criticism of the Average Boy is found by counting the number of times "Ordinary" differs from "Ideal." When this score is high, the individual rates the group below the ideal frequently, and hence tends to criticize it. When the score is low, the average boy is rated as being close to the ideal.

Feeling of Difference is found by counting the number of times "Self" is marked differently from "Ordinary." When this is high, the boy feels himself to be different from most boys; when low, he feels himself to be an average or typical boy.

Superiority and Inferiority constitute two separate scores. When "Self" is rated nearer to the "Ideal" than "Ordinary" is, one point is credited to Superiority. When "Ordinary" is rated nearer the "Ideal" than "Self" is, one point is credited to Inferiority. These two scores should be interpreted in conjunction with one another.

Deviation from the Accepted Idea of the Right is found by counting the total number of items in which 20 per cent or less of the group have marked "Ideal" as in the same manner as the individual being scored. This requires determining the frequency of the distribution of replies to "Ideal" for the group taking the test and marking on a fresh sheet those replies made by 20 per cent or less of the group. High scores on this factor indicate that the individual diverges from the ideal of the group in which he is tested; low scores indicate that he approaches the common ideal.

Social Insight is measured by counting the number of items in which 20 per cent or less of the group answer "Self" as the individual himself marks "Ordinary" and subtracting the result from 51. To obtain the score, a frequency distribution of the "Self" scores made by the group is necessary. "High Social Insight" scores indicate that the boy believes that boys feel very much as they actually do say they feel; low scores indicate that the individual whose paper is being scored misjudges by his answers to "Ordinary" how other boys feel as indicated by their answers to "Self."

These scores are highly reliable, as is shown in the following

Diagnosing Personality and Conduct

table of average reliability coefficients from four groups with from 50 to 136 boys in each group.

TABLE 4

RELIABILITY COEFFICIENTS FROM PERSONAL ATTITUDES TEST (SELF-ORDINARY IDEAL RATING)

(from Sweet, 100, p. 48)

	<i>Average reliability coefficient</i>
Self-criticism914
Criticism of others939
Feeling of difference935
Superiority936
Inferiority785
Deviation from accepted idea of right86
Social insight87

Certain of these scores, particularly the criticism of self, criticism of others, and a combination of superiority-inferiority scores, showed promise of being diagnostic of the presence of personality and behavior problems when checked against psychiatric case studies and the ratings of social workers.

Comparison of Rating Methods

Various factors should be considered in determining which method of rating is best. Objectivity and reliability play an important part. Convenience, ease in handling, comprehensibility of method, and freedom from variation are all matters to be considered.

In general it may be claimed that the different methods are equally objective. Objectivity in rating depends on the method by which the rater makes his observations, and draws his inferences and on the cleverness with which the positions on the scale are defined rather than on the way in which the judgments are recorded. Barrett (7) assures us that the order-of-merit method and the method of paired comparisons are equally good. The present writer (101) has determined that rating and ranking are equally reliable under ordinary conditions in school. The choice of method in recording one's judgments depends on other factors than reliability.

One advantage claimed for ranking, paired comparisons, and the graphic method over the method of assigning definite scale values

is that discriminations may be made as fine as are desired, at least up to a certain degree. However, it is possible to push the matter of fineness of discrimination too far. One who attempts to rank thirty or forty individuals in some trait finds that, after ranking four or five outstanding individuals who can readily be ranked, the rest of the cases seem much alike, and accurate rankings become very difficult to assign. Sometimes ranks are the result of forced discriminations which are not actually felt by those who have been required to make them. For this reason ranking becomes irksome when the number of cases to be ranked is large. Rating, particularly on a graphic scale, is more pleasant.

Some consider the method of ranking easier to explain and easier to understand than the method of rating. Ranking usually gives a false notion concerning the form of the distribution of the trait being rated, because ranking yields a rectangular distribution, whereas seldom if ever do traits for any group assume a rectangular distribution. The most likely assumption we can make about any distribution is that it approximates the normal probability curve. In ranking, the end cases are too close together and the middle cases are too far apart. However, by statistical methods the ranks may be turned into numbers which yield a distribution having the shape of the normal distribution. The method for doing this is explained on page 91.

Knight* objects to rating on the ground that rating may be biased, i.e., be either too low or too high—usually too high. He says, "Judges who are presumably equally competent rate the same person as almost wholly honest, half honest, rather dishonest, or an out-and-out cheat. This disagreement is the product of legitimate variations in knowledge about the person and of varying standards of honesty held by the judges themselves. The way to avoid this pitfall is not to use a rating involving concepts of virtues, but a system of ranking persons according to their relative merits. The question should be not, 'How honest is he?' but 'Is he more or less honest than this person or that?'"

One teacher might give a class an average rating of 4 on a rating scale in honesty and another teacher give the same class an average rating of 2, and yet the two sets of ratings might correlate perfectly. This tendency to rate too high or too low may be

* Knight, F. B., "Analysis of Teaching and Teachers," *Jour. of Educational Research*, 10:224 (Oct., 1924).

corrected (see p. 97), so that rating need not be discarded for this reason.

Conklin and Sutherland (24) have pointed out that rating is better for getting an immediate affective impression, such as the humor value of jokes. In rating, the values once assigned are not easily changed. In ranking, on the other hand, it is easy to shift the position of a card in the pack and thus change the original impression. Using the same argument, rankings should be better for recording mature or pondered judgment.

Cardinal Points to Be Observed in Rating

The number of scale-divisions in the rating scale. The two rating scales by Galton and Pearson given as examples contained nine and seven scale-divisions respectively. Compare these with the data of the following table, from Boyce (11), showing the number of scale-divisions in fifty-four teacher-rating schemes which he examined.

TABLE 5
DISTRIBUTION SHOWING NUMBER OF SCALE-DIVISIONS IN FIFTY-FOUR
RATING SCALES *

(from Boyce, 11, p. 20)

<i>Number of scale divisions in rating scale</i>	<i>Frequency</i>
2	2
3	11
4	24
5	12
6	2
7	1

54

* From Boyce, A. C., "Methods for Measuring Teachers' Efficiency," *14th Yearbook, National Society for the Study of Education* (1915), Part II. Reprinted by permission of the Society.

In this matter there are forces pulling in two directions. On the one hand, for accurate measurement we want as fine discrimination as possible and hence as many scale-divisions as possible. One of the advantages claimed for the graphic method is that it permits as fine discrimination as may be desired. In the other direction there are factors of economy in effort. Naturally it takes more effort to sort persons into five divisions than into three. There

is an upper limit to fineness of discrimination and it is possible to have one's scale divided too finely for feasibility in practice. We do not need a yardstick with inches divided into tenths if the scale must be read at a distance of fifty feet. Just as we may have a scale divided finer than the eye can read, so it is possible to devise a rating scale with more classes than any one can discriminate. The potential power of discrimination with the graphic scale may be greatly overestimated, since one's powers of judgment will fail long before the divisibility of the scale is exhausted.

The present writer (102) attempted to determine the optimum number of classes for a rating scale. He showed that the decisive point at issue is the loss of reliability which one will accept due to coarseness of the scale. Setting up the arbitrary standard that a loss in reliability will be accepted equivalent to the change from a reliability of .91 to one of .90, the author computed that seven is the optimum number of classes for rating human traits. With less than seven classes the coarseness of the rating led to an appreciable loss in reliability greater than his arbitrary standard of allowable loss. More than seven classes yielded such a small increment in reliability above what would be obtained with seven classes that the attempt to make the finer discrimination turned out not to be worth while.

Although seven classes is an average optimum number for a rating scale, conditions are often such that more or fewer classes are justified. If the trait is an obscure one such as impulsiveness, or if the raters are untrained or take only a moderate interest in the task, rating scales with more than three or four classes are inexpedient. If a scale consists of several items which are cumulated to make the score on the scale, it is not necessary to rate each item as accurately as though each stood alone.

On the other hand, if one is rating fairly objective functions such as difficulty of words in spelling, or if the raters are enthusiastic and trained, more than seven are justified. In general, table 6 on the next page may serve as a guide in determining the number of classes.

It is a curious fact that there seems to be a natural reluctance to use seven classes in rating. In school marks, for instance, it is customary to use a five-letter system. However, if we undertake to measure at all, we should measure as accurately as possible. The more accurately we rate, the less injustice we shall do to the

TABLE 6

NUMBER OF CLASSES FOR RATING SCALES YIELDING DIFFERENT RELIABILITIES

<i>Obtained reliability</i>	<i>Number of classes desirable</i>
.95	18
.90	14
.80	11
.70	9
.60	7
.50	6
.40	5
.30	4
.20	3
.10	2

individuals being rated. Accordingly seven classes should be preferred to five classes in rating human traits.

The number of individuals in each class. Ratings can be studied by bearing in mind the approximate numbers of individuals who as a rule should fall in each class. This does not in the least bind the rater, but it does serve as a guide and will help to make ratings by different examiners comparable. Of the various inevitable assumptions which must be made in ratings, the most probable assumption is that the group is a typical one and that consequently the members of the group distribute themselves normally with respect to the trait. Table 7 on the next page, derived from tables of the probability integral, may be used for guidance in determining the number of individuals who should fall in each class. A consistent use of these figures will do much toward increasing the value of ratings and making the ratings by different observers comparable.

Rating one quality at a time. One of the two methods of rating a group is to rate an individual through all the traits on the sheet, and then pass on to the next individual. Another way is to rate all individuals through on a single trait before passing on to the next trait. It is generally conceded to be preferable to rate one trait at a time, because a peculiar phenomenon occurs as one rates an individual on several traits. Either through mental inertia on the part of the rater or because an individual usually seems to present a sort of uniform level on all sides of his personality to the observer, it has been noted that the ratings of an individual tend to be monotonously alike. If the individual is high in one trait he is likely to be high in all; if low in one, low in all. It is demonstrably

TABLE 7

PERCENTAGE DISTRIBUTIONS FOR GROUPS OF DIFFERENT SIZES AND FOR SCALES
HAVING DIFFERENT NUMBERS OF CLASSES

CLASS OF 40 (RANGE 5 S.D.)				
3 classes	4 classes	5 classes	6 classes	7 classes
20%	11%	7%	5%	4%
60	39	24	15	10
20	39	38	30	22
	11	24	30	28
		7	15	22
			5	10
				4
CLASS OF 185 (RANGE 6 S.D.)				
3 classes	4 classes	5 classes	6 classes	7 classes
16%	7%	4%	2%	2%
68	43	24	14	8
16	43	44	34	23
	7	24	34	34
		4	14	23
			2	8
				2

true that our ratings for any one trait are colored by the general impression of the individual which we have formed. This tendency to uniformity does not appear where one trait is rated through at a time, for in the latter case there is a direct comparison of several individuals with respect to a single trait, with results which are distinctly superior to comparing several traits in a single individual.

Selecting the qualities to be rated. If one is rating a narrow and definite habit or attitude or trait, this problem does not arise. If, however, one is rating for the purpose of estimating efficiency or the value of individuals for a certain kind of work, or if one is rating pupils on their ability to do school work, careful attention must be given to the qualities to be rated. Some qualities are quite unimportant, at least for the purpose for which the rating is being done. A clerical worker need not be rated, for instance, on his temperament, nor a teacher on thrift, nor a school pupil on reverence. Some qualities are relevant, others are not. It is possible to rate "general efficiency," but there is no evidence to show that this

method does not yield results inferior to a more analytical sort of rating. It seems reasonable to suppose that a summation of the ratings of the analyzed components of "general efficiency" should be more valuable than a blanket rating. Separating the rating into separate qualities helps to standardize the rating. If the rating is on general efficiency, then such factors as a disfigured face or a brusque manner may overweigh or prejudice the estimate. If such factors as intelligence, industry, skill, and coöperativeness are the important factors in determining a man's value, rating on these factors would result in more representative estimates than would be obtained if each rater used his own judgment as to what factors were important. Ratings on several factors may not yield results which are much more reliable than rating on a single factor, but the former method should help eliminate personal bias.

The analysis of man's personality into separate traits is fraught with difficulties. One never knows how much one is drawing on his imagination in making such an analysis. He does know whether a careful analysis into qualities would yield the same results as an experimental determination of the analysis. The most approved method of choosing the items for a rating scale is to rate the group of persons on whom the scale is finally to be used both on general efficiency and also on a wide variety of separate qualities. Those qualities correlating most highly with the criterion (which should be an independent measure of general efficiency) would be the ones to be used in the rating scale.*

Care must be exercised in selecting the items for a rating scale that they do not overlap. Often apparently different qualities turn out to be practically identical. For instance, school attitude, effort, and industry are identical to all intents and purposes. Moreover, one should be careful not to permit obscure designations of the qualities to be rated. Persons with inadequate psychological training often make fantastic analyses of personality.

Weighting qualities to be rated. In some of the older work with rating, when score cards were more frequently used, more attention was paid to weighting the qualities to be rated. Naturally every one would agree that in the case of a clerical worker, speed

* Strictly speaking, the intercorrelations between all traits should be found and regression weights computed which will make the best weighted composite and hence the highest correlation with the criterion; then those yielding the highest regression weights for predicting the criterion would be chosen for the scale.

and diligence and intelligence are more important than sense of humor, taste in dress, or politeness. A properly constructed score card is a reminder of relative values. In determining the weights to be given to the items in a score card, it is customary to get composite judgments. Usually 100 or 1,000 points are decided upon for the score card, whereupon the list of qualities is sent to competent and interested persons who are requested to distribute the 100 or 1,000 points among the items. With very slight adjustment, the average of the allotments may then be used as the weights.

Another and more difficult yet more scientific method of assigning weights is to rate a number of persons both by a blanket rating of efficiency and by the items on the card. Correlations with the criterion and intercorrelations between the items may be used to determine the regression weights to be used as the weights of the items in the scale.

Probably in ordinary work where ratings are as unreliable as they usually are, the matter of weighting is relatively unimportant, though it should be remembered that the matter of weighting cannot be ignored and that in a true sense the items are always weighted, since they are weighted equally if nothing else. But because the method of obtaining regression weights is difficult and expensive, we assure ourselves, with probable justification, that some more approximate method of determining weights is usually quite as satisfactory.

Number of items in a rating scale. A further issue that must be decided is the number of items to include in the scale. If one is making a score card for the purpose of analyzing school buildings, the extent of the analysis is determined partly by the functions of the different parts of a building. But even in this instance judgment must be used as to how detailed the rating is to be. In general we do not want to make the rating sheet too detailed, with the resulting increase in the labor of filling it out. The issue reduces itself to the matter of the gain in reliability and validity that accrues from the summation of ratings over and above what would be obtained from a blanket rating. Although there is no precise evidence on this issue, the present writer's opinion is that the value of the analysis is usually achieved quickly, so that in general a rating scale need not contain over three to five items. The increased accuracy which is gained from more detailed ratings is probably not very great.

Items in a rating scale must be observable. In selecting qualities to be rated, it is important to recognize that they should be qualities on which the observer has an opportunity to obtain data. For instance, an interviewer might be in an excellent position to give a rating on appearance, but would not be able to form any judgment as to coöperation or industry. In choosing items for a teacher-rating card to be used by a supervisor, it is necessary to make sure that the items are such as can be observed during a class-room visit. One could easily observe and form a judgment as to appearance, voice, or use of English during a supervisory visit, but one might obtain no evidence permitting an evaluation of the growth of pupils with respect to subject-matter, definiteness and clearness of aims, optimism or extra-mural interests. It is astonishing how often rating scales are drawn up asking judgments for which the rater is not in a position to gather evidence.

Definition of qualities to be rated. Particular attention must be paid to the definition of the items in the scale. On this hinges much of the success or failure of ratings in general. One of the most potent factors causing unreliability of ratings is ambiguity in meaning of the items on the scale. Every rating is a judgment; judgments depend not only on observation, but also on the interpretation according to certain standards of what is observed. *Intelligence* is a good example of a term which has an ambiguous meaning. To one man it will mean "adaptability," to a variety of other men it may signify "abstractness," or "agility," or "ability to learn," or "possession of general information," or "constructiveness," or "ability to put things together." *Character* is another ambiguous term often used on recommendation forms. It may mean "moral character," or "trustworthiness," or "sexual purity," or "prompt payment of bills," or "coöperativeness," or even "general value to society." *School attitude* is still another ambiguous term interpreted as anything from "industry" to "docility." In every rating scale the items should be defined in some way. There are several possible ways of doing this. One, perhaps the least satisfactory, is to give synonyms of the original term. Another is a short paragraph amplifying the descriptive title. Another method is to ask a question which not only limits the meaning of the term but somehow helps the rater to see the problem of rating more clearly. Still another method is that of describing a person who possesses the trait under consideration.

Objective vs. subjective definition. Paterson (79, p. 83) emphasizes the point that the trait should be defined objectively rather than subjectively. A subjective definition is in terms of the man's personality and traits; an objective definition is in terms of what the man does and particularly what effect the man has on the person or things with whom he comes in contact. Paterson gives examples of objective and subjective definitions.

Leadership

Subjective definition: Rate this executive's force, self-reliance, decisiveness, tact, ability to inspire men and to command their obedience, loyalty and coöperation.

Objective definition: Rate this executive according to the success he has shown in developing a loyal and effective organization by administering justice, inspiring confidence, and winning the coöperation of his subordinates.

Appearance

Subjective definition: Personal attractiveness, cleanliness, neatness, dress.

Objective definition: Consider how favorably he impresses his men by his physique, bearing, and manner.

Many writers have stressed the point that the definition of items on the rating scale should be objective. Abstract traits standing alone may be highly subjective, but when related to a definite situation they may acquire objectivity. For instance, *independent*, standing alone, might be useless in describing personality, yet might be very significant when used in connection with specific situations such as scientific research. Hughes (54) pleads for a behavioristic definition, by which he means a definition more in terms of what a person does and what products or effects result from his activity than in mere descriptive terms. Paterson (82) urges that the rater keep in mind past or present accomplishments of the individual being rated. Judgment should be reached, not on the basis of hoped-for achievement or possible development but on the past record of the candidate as reported or observed.

Not only the trait itself, but the degrees of the trait should be clearly and sharply defined. Miner (78), an early student of rating methods, stated that the difficulty of defining verbally the de-

degrees of a trait was so great that it was better to avoid qualitative terms and substitute relative or quantitative terms, as by a scale with units equal to fifths of the group. This emphasis on a direct comparison and grouping of individuals is sound, but this method belongs primarily in the hands of the skilled rater. Later writers plead for clearer definitions of the degrees of a trait. It has been suggested that the words *average*, *fairly*, *very*, *exceedingly*, and *extremely* should be avoided in favor of more descriptive adjectives. Freyd (32) gives illustrations: *fastidious*, in place of *extremely neat*; *slovenly*, in place of *very careless in dress*. On the other hand difficulty of vocabulary should be considered, with exclusions of unusual words even though they are highly descriptive and meaningful. Other things being equal, the more common the descriptive adjective used the better. Probably the best results are obtained from making the definitions of the degrees of a trait objective.

Statistical Treatment of Paired Estimates and Rankings

Ratings are usually treated at their face value. The steps on a rating scale which has been assembled hastily are only approximately equal. However the errors due to unreliability of rating are greater than errors due to carelessness of scaling, so that the latter can for practical purposes be disregarded.

The method of paired estimates needs statistical treatment before the several individuals judged may be assigned a numerical rating. As the judgments are made in the methods of paired estimates, each individual is stated to be better or worse than each other individual in turn. These judgments must then be manipulated so that each individual acquires a position along a linear scale.

Thorndike (109) was the first to propose a definite technique for combining order-of-merit judgments. His method essentially was first to place the individuals in their order by inspection, then to determine the true difference in position between each pair by finding the percentage of judges who rated the former individual as lower than the latter, and finally computing the amount of difference represented by the percentage. If individuals were transposed, i.e., wrongly placed by the original inspection, they must

be shifted, and the calculations concerning each individual affected by the shift recomputed. Thorndike directs (106, p. 198): "In connection with this amendment it will be useful to compute also the differences from the two or three next neighbors in the new order. For if an individual is rightly placed, he will not only be below the one placed above him in the order, but also below the two or three next above him. A process of trial and adjustment will be more economical than a rigid procedure in this work." . . . (109, p. 204): "Perhaps some single system of over weighting, the next neighbor comparison would be advisable. On the whole, however, we have in this technique an economical and fairly reliable means of combining partial orderings." It will be noted that the method proposed by Thorndike is by no means an exact one.

After the differences are recorded in percentages, these percentages may be turned into absolute amounts. A unit of absolute difference is defined as that corresponding to a difference noticed by 75 per cent of the judges. If desired, an arbitrary scale may be assumed and individuals placed on the scale by adding the absolute amounts of difference.

TABLE 8

THE AMOUNTS OF DIFFERENCE ($x - y$) CORRESPONDING TO GIVEN PERCENTAGES OF JUDGMENTS THAT $x > y$

% r = the percentage of judgments that $x > y$

$\frac{\Delta}{P.E.} = x - y$, in multiples of the difference such that % r is 75

% r	$\frac{\Delta}{P.E.}$	% r	$\frac{\Delta}{P.E.}$	% r	$\frac{\Delta}{P.E.}$	% r	$\frac{\Delta}{P.E.}$	% r	$\frac{\Delta}{P.E.}$
50	.00	60	.38	70	.78	80	1.25	90	1.90
51	.04	61	.41	71	.82	81	1.30	91	1.99
52	.07	62	.45	72	.86	82	1.36	92	2.08
53	.11	63	.49	73	.91	83	1.41	93	2.19
54	.15	64	.53	74	.95	84	1.47	94	2.31
55	.19	65	.57	75	1.00	85	1.54	95	2.44
56	.22	66	.61	76	1.05	86	1.60	96	2.60
57	.26	67	.65	77	1.10	87	1.67	97	2.79
58	.30	68	.69	78	1.14	88	1.74	98	3.05
59	.34	69	.74	79	1.20	89	1.82	99	3.45

Thurstone (III) has demonstrated a method whereby items judged by paired comparisons may be compared without resort to approximations or estimates. This method is given below because it is probably the best available way of reducing paired comparisons.

88 Diagnosing Personality and Conduct

Step 1. Let there be five variables, a , b , c , d , and e , each of which has been compared with every other by a number of judges. The following table gives the percentages of judges who believe an item in any column is less valuable than the items listed at the left.

TABLE 9
PER CENT OF JUDGES WHO BELIEVE ONE ITEM TO BE LESS
VALUABLE THAN ANOTHER

	a	b	c	d	e
a323	.338	.211	.128
b677		.415	.242	.172
c662	.585		.260	.136
d789	.757	.740		.379
e872	.828	.864	.621	
Sum	3.000	2.493	2.357	1.334	.815

That is, item c (looking in column c) is judged to be less valuable than item b (looking in line b) by 41.5 per cent of the judges. Likewise item b is judged to be less valuable than item c by 58.5 per cent of the judges.

It will be noted that the sums of the columns of percentages indicate that the variables are in rank order. If they are not in rank order, they should be placed in rank order in proceeding to step 2.

Step 2. Turn these percentages into their equivalent deviations in terms of their respective standard errors of observation. The Kelley-Wood Table of the Normal Probability Integral* is used. The percentages are read in columns headed p or q , while the deviations in terms of the respective standard errors of observation are read on the same line in column z .

TABLE 10
PERCENTAGES OF TABLE 9 TRANSMUTED INTO CORRESPONDING
STANDARD DEVIATION VALUES

	a	b	c	d	e
a		-.46	-.42	-.80	-1.14
b46		-.21	-.70	-.95
c42	.21		-.64	-1.10
d80	.70	.64		-.31
e	1.14	.95	1.10	.31	
	2.82	1.40	1.11	-1.83	-3.50

* Kelley, T. L., *Statistical Methods* (The Macmillan Company, 1923), p. 371.

Step 3. The deviations are then added algebraically.

Step 4. The "scale" separation of adjacent columns is computed by the following formula.

$$\text{where } S_1 - S_2 = \frac{\sqrt{2}}{n} [\Sigma x_{1k} - \Sigma x_{2k}]$$

$S_1 - S_2$ = scale separation

n = number of values in a column

Σx_{1k} = sum of first column of pair

Σx_{2k} = sum of second column of pair

For instance, in computing the scale separation of *a* and *b* one obtains:

$$S_1 - S_2 = \frac{\sqrt{2}}{5} [2.82 - 1.40] = .402$$

These values become

$$S_{a-b} = .402$$

$$S_{b-c} = .082$$

$$S_{c-d} = .832$$

$$S_{d-e} = .472$$

Step 5. These may be placed more conveniently in a scale having only positive values by setting the scale value of *e* arbitrarily at .000 and adding the scale differences between successive variables.

Scale values

<i>a</i>	1.788
<i>b</i>	1.386
<i>c</i>	1.304
<i>d</i>472
<i>e</i>000

Thurstone recommends that percentages in step 1 greater than 97 or less than 3 be omitted from the table as being too unreliable. The corresponding standard errors of observation must then be omitted from the table in step 2. This necessitates a slight change in the procedure in steps 3 and 4. In step 3 only those values will be added for which there are corresponding values in the adjacent column. This may mean that any one column may have to be added twice, in order to provide corresponding scores for the columns on its left and on its right. In step 4, *n* must always be the exact number of values added in the column.

TABLE OF STANDARD DEVIATIONS TO BE ASSIGNED FOR RANK POSITIONS

[illegible][illegible]

Ranks yield a peculiar kind of distribution. In ranking, every individual is one step away from the next person. The distribution formed is said to be *rectangular* and is quite unlike the form of distribution obtained from testing with a scale in equal units.

DISTRIBUTION OF RANKS

.1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
----	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----

When we want to compare ranked data with data obtained from a linear scale, as scores on tests, ratings, etc., the ranks must be transmuted into equivalent scores. The best method is to transmute the ranks into standard deviation units. Tables have been devised for this by Ream (87)* and Chu (20), but these tables are inaccurate. Table 11 on page 90 has been computed afresh from the Kelley-Wood Tables of the Normal Probability Integral.

Another method of transmuting rankings into units of amount was devised by Hull. This method consists first of turning each rank into a per cent, using the formula $\frac{100(R-.5)}{N}$ where N is the total number ranked, and then estimating the score from Table 13 on page 92.

The following example shows how the method, including the table, is used.

Given eight pupils ranked in order of industry, to transmute these rankings into units of amount on a linear scale:

Pupil A, who has a rank of 1, has a per cent position of $\frac{100(1-.5)}{8} = 6.25$. His score from Table 13 is nearest 8.0.

TABLE 12

THE RANKING OF EIGHT PUPILS TURNED INTO UNITS OF AMOUNT

Pupil	Rank	Per cent position	Score
A	1	6.25	8.0
B	2	18.75	6.7
C	3	31.25	6.0
D	4	43.75	5.3
E	5	56.25	4.7
F	6	68.75	4.0
G	7	81.25	3.3
H	8	93.75	2.0

TABLE 13

TABLE FOR TRANSMUTING RANKINGS INTO UNITS OF AMOUNT
(from Hull)*

Per cent	Score	Per cent	Score	Per cent	Score
.09	9.9	22.32	6.5	83.31	3.1
.20	9.8	23.88	6.4	84.56	3.0
.32	9.7	25.48	6.3	85.75	2.9
.45	9.6	27.15	6.2	86.89	2.8
.61	9.5	28.86	6.1	87.96	2.7
.78	9.4	30.61	6.0	88.97	2.6
.97	9.3	32.42	5.9	89.94	2.5
1.18	9.2	34.25	5.8	90.83	2.4
1.42	9.1	36.15	5.7	91.67	2.3
1.68	9.0	38.06	5.6	92.45	2.2
1.96	8.9	40.01	5.5	93.19	2.1
2.28	8.8	41.97	5.4	93.86	2.0
2.63	8.7	43.97	5.3	94.49	1.9
3.01	8.6	45.97	5.2	95.08	1.8
3.43	8.5	47.98	5.1	95.62	1.7
3.89	8.4	50.00	5.0	96.11	1.6
4.38	8.3	52.02	4.9	96.57	1.5
4.92	8.2	54.03	4.8	96.99	1.4
5.51	8.1	56.03	4.7	97.37	1.3
6.14	8.0	58.03	4.6	97.72	1.2
6.81	7.9	59.99	4.5	98.04	1.1
7.55	7.8	61.94	4.4	98.32	1.0
8.33	7.7	63.85	4.3	98.58	.9
9.17	7.6	65.75	4.2	98.82	.8
10.06	7.5	67.48	4.1	99.03	.7
11.03	7.4	69.39	4.0	99.22	.6
12.04	7.3	71.14	3.9	99.39	.5
13.11	7.2	72.85	3.8	99.55	.4
14.25	7.1	74.52	3.7	99.68	.3
15.44	7.0	76.12	3.6	99.80	.2
16.69	6.9	77.68	3.5	99.91	.1
18.01	6.8	79.17	3.4	100.00	0
19.39	6.7	80.61	3.3		
20.93	6.6	81.99	3.2		

It is often necessary to combine rankings or ratings where no judge has ranked or rated all of the individuals. For instance, if each member of a school faculty were asked to rate the other members, it is obvious that in some cases acquaintance would be so slight as to render an accurate judgment impossible. It is the usual practice, therefore, to ask one person to rate or rank in order merely the ten or a dozen persons he knows best. As a result, the ratings are incomplete, i.e., do not include the entire group.

* From Hull's *Aptitude Testing*, p. 387. Copyright 1928 by World Book Company, Publishers, Yonkers-on-Hudson, New York.

Garrett (36, p. 171) has studied the matter of combining incomplete rankings (or ratings). He considered four methods: (a) turning ranks into standard deviation units (as explained on p. 91), (b) using a simple average of ranks, (c) turning ranks into percentiles (as explained on p. 91), and (d) making a comparison according to the method proposed by Thorndike or Thurstone (see pp. 86-89). His conclusions are:

"1. The final order of amount obtained from incomplete judgment lists by the S. D., Average, Percentile, or Comparison methods tallies very closely with the best 'standard' or 'true' order. The more numerous the partial lists, and the longer the lists (other things being equal), the closer the agreement of the final order from the partial lists and the final order from complete lists.

"2. The order of amount obtained from scattered and sparse data may be very inaccurate when judged by the order obtained from complete lists. With such data no method gives accurate results.

"3. Judged from the standpoints of simplicity and time required, either the Average method or the Percentile method is superior to the other methods. As far as accuracy is concerned, the Average method is certainly as good as the other three, if not better.

"4. When the variability of the individuals or things rated is desired, either the S. D. or the Percentile method may be used to advantage. Of the two, the Percentile method is the more simple as it avoids the use of negative quantities."

In combining two sets of ratings the two factors of *importance* and *variability* must be considered and allowed for. If ratings have been turned into standard deviation units or percentiles, the variability of different rating scales may be taken as equal and no allowance need be made for this factor when combining.

Reliability of Ratings

The reliability of ratings has been found to be variable and disappointingly low; indeed, it has been found by various workers to be so low as to cast grave doubt on the value of rating as a method for gathering trustworthy data.

The reliability of personal ratings is usually measured by the correlation between ratings by two comparable judges.*

* Shen (94, p. 232) has pointed out that this procedure for determining the reliability of ratings really leads nowhere, because it is next to impossible to

Miner (78) reports a reliability of .54 when sixty-four college seniors were rated by instructors for general ability and .57 when thirty-six students were similarly rated. Hughes (55) reports an average reliability coefficient of .56 when 253 high school students were rated on twelve traits and an average reliability of .63 when seventy-nine students were rated on twelve traits. Shen (94), in a careful study, reports an average reliability coefficient of .55 for thirteen judges (presumably college students) rating themselves and each other on eight traits. Webb (120) obtained an average reliability of .55 when ratings were made on forty-five traits of 194 school-boys rated in groups of twenty (average age twelve); and another average reliability of .55 when 140 young men in groups of about thirty-five (average age twenty-one) were rated on twenty-eight traits. Webb admits having rejected fifteen pairs of estimates out of 112 for the boys, and sixty-three out of 445 for the students. Using his complete data the average reliabilities are boys .49 and students .47. Waite (117) found reliabilities of .47 and .50 when 1,405 and 2,018 pairs of judgments were obtained on school-children. Hayes and Paterson (43) state in a report on the graphic rating scale that correlations between eight supervisors were as a rule over .65. Gallup (35) reports the use of a graphic rating scale, which yielded a correlation with the educational director's ratings of .66 in estimating the success of retail salespeople.

Kornhauser (69, p. 120) reports on ratings of speed, accuracy, and general value for a group of specialized office workers. "The ratings by the three supervisors agreed very closely with one another even though they were submitted quite independently. The correlations between the ratings by each supervisor and those by each of the other supervisors also fall between 0.70 and 0.90." It is unfortunate that Kornhauser did not report more fully the

obtain two persons who judge with equal reliability or, in other words, who are comparable judges. "Hence a correlation between two judges is a very crude approximation to the reliability of either. The reliability of a judge thus crudely evaluated often varies considerably according to the judge with whom he happens to be correlated." Shen has derived a formula for the elimination of this systematic error in the determination of reliability which he calls correction for pulling, based on the proposition "the correlation between two series of ratings on the same trait, independent in errors of each other, is equal to the geometric mean of their true reliabilities." Practically all reliabilities reported in psychological literature, however, are obtained by merely correlating the ratings of one judge against those of another.

conditions under which he obtained such high reliability coefficients.

Furfey (34) reports that analyzing a trait (developmental age) into several sub-traits, having the judge rate all these sub-traits separately on seventy-five boys, and then combining these separate ratings into a final score gives the high reliability coefficient of .888.

A reliability coefficient of .55 can be said to be typical for rating personality traits by ordinary judgment methods. Some traits yield higher reliability coefficients, others lower. It is easy to fall short of even this average figure of .55 if the raters are careless, if the traits are loosely defined, if acquaintance with those being rated is slight, or if there has been inadequate observation. There is no evidence that either the man-to-man method or the graphic rating scale yields higher reliabilities when only one characteristic has been rated. However, when several traits are rated independently and their results combined, the reliability may be raised to much higher figures. Indeed, if a trait is analyzed into sub-traits which have been found by analysis to correlate highly with the trait, and ratings on the sub-traits by an individual are combined into one rating, the result may have a reliability so much higher that it is only slightly under what might be predicted by the Spearman-Brown formula from a single rating. On the other hand, if the analysis into sub-traits is subjective, the composite rating may be no better than the blanket rating.

The conclusion drawn from facts concerning the reliability of ratings is that in general the rating by a single judge is too unreliable to be useful. Rugg (90) has set up the standard gage that human character can be rated accurately enough for practical purposes in education only when the rating is the *average of three independent ratings*. The number of independent ratings which should be obtained may be determined by means of the Spearman-Brown formula.* The accompanying table enables one to know how many independent ratings are needed to obtain certain desired reliabilities. In one column is given the number of ratings necessary to obtain a reliability of .82—an average figure for a forty-five-minute standardized test of achievement in high school. In the other column is given the number of ratings necessary to

* Chu (20) and Remmers, Shock, and Kelly (89) give data to show that ratings do follow the law expressed in the Spearman-Brown formula.

obtain a reliability of .90—good for satisfactory individual diagnosis.

TABLE 14

NUMBER OF RATINGS NECESSARY TO OBTAIN A SPECIFIED RELIABILITY
FOR VARIOUS TRAITS

	r_{11}	NUMBER OF RATINGS TO OBTAIN A RELIABILITY OF	
		.82	.90
Average trait55	4	8
Scholarship71	2	4
Leadership68	3	5
Intellectual quickness68	3	5
Intellectual profoundness64	3	6
Memory55	4	8
Persistence40	7	14
Adaptability37	8	16
Impulsiveness34	9	18

As a matter of fact, then, Rugg erred on the side of leniency in requiring only three independent ratings. The above table shows *eight* as the average number of independent ratings one should obtain if the ratings are to be individually diagnostic.

Tendency to Rate Too High

Unreliability is not the only charge to be made against rating. The tendency to rate too high or to vary the standard of rating introduces other serious errors. Two raters may show high agreement so far as the reliability coefficient indicates, and yet may so differ in their level or standard of judgment that their ratings of the same person differ considerably. This lack of uniformity in standards is due very largely to differences in the interpretation of the descriptive adjectives which define the steps of the scale. Boyce (11) says, "A teacher who ranks Excellent in one officer's mind is found to rank only Good or perhaps Medium in the rating of another judge who is more critical and less easily satisfied."

It has been found that ratings tend, on the average, to be too high, often so high that the lower end of the scale may not be used at all. Evidently we have been overgenerous in judging if only the upper half of a rating scale has been used when it

was expected that the ratings would be distributed normally. Terman calls this the generosity factor. He believes "that we always tend to overrate those we like" and that "when we are rating those we dislike, the generosity factor probably operates negatively."

As a correction for this error, some graphic rating scales have been so constructed that most of the descriptive terms represent above average in the trait. Instead of

bad	poor	average	good	excellent
-----	------	---------	------	-----------

a scale will read

poor	fair	good	very good	excellent
------	------	------	-----------	-----------

thereby taking advantage of the frailty of the human mind in judging others.

Several suggestions have been made for correcting ratings and thereby reducing them to a common basis. Paterson (82, p. 91) advises the following method: "The reports made by a given supervisor or foreman are assembled and the 'total scores' are arranged in a frequency distribution from high to low. This distribution is then divided into five parts so that the highest 10 per cent of the total scores are given a final letter rating of A, the next 20 per cent are given a final letter rating of B, the next 40 per cent are given a C rating, the next 20 per cent are given D, and the lowest 10 per cent are given E. The limiting points in the total score are noted and a 'key to final ratings' is prepared whereby future ratings made by this supervisor may readily be converted into final ratings. This procedure converts the actual ratings into relative ratings and is designed to do away with the error which otherwise arises because some supervisors rate too high and other supervisors rate too low." This correction is based on the assumption that every supervisor is rating a group which distributes itself normally in a trait and that every group has the same distribution as every other group. Of course this is never exactly true, but without further knowledge it is the best assumption that can be made. Indeed, it is probably a truer assumption in any given case than that certain groups are select.

Two other suggestions have been made for eliminating the effects of differences in standards. As mentioned before, Knight (63) suggests ranking instead of rating, for then individuals are

compared with each other, though nothing is said concerning the absolute level of the group as a whole, nor that of any individual.

The other suggestion (101) is to incorporate at the top of every graphic rating scale a series of percentages (as in Rating Chart B on page 64) which will suggest the number to go in each compartment. This method will reduce but will not eliminate differences in standards in rating.

Factors Influencing the Reliability of Judgments

There are many factors contributing to the unreliability of ratings. An understanding of these factors may help to lessen or eliminate the effect of some of them and may lead workers to appreciate the difficulties involved in obtaining reliable ratings. In the first place it seems to be easier to rate some people than others. Secondly, judges differ in their ability to give reliable ratings. In fact, the same judge at different times and under different circumstances will differ in the reliability of his judgments. Thirdly, the traits or acts being rated differ in the degree to which they can be reliably rated.

Differences in persons being rated. Norsworthy has demonstrated that we agree in our judgments about some people better than about other people. Seven college girls rated ten members of a sorority. From the figures, "it would seem that among ten girls who know each other well there may be twice as much difference of opinion about some one member of the group as about some other." What factors are involved here is obscure.

The effect of acquaintance on the reliability of rating has not been very thoroughly studied. Cleeton and Knight (21) found that there was practically no correlation between the ratings of casual observers and close acquaintances on the same individuals. On the other hand Landis (72) reports that the reliability of ratings does not differ whether made by intimate associates or by general acquaintances. Shen (93), who has made the most analytical study of the effect of acquaintance on ratings reports that the average error of ratings is not affected by the degree of acquaintance. The "casual observers" of Cleeton and Knight were members of an audience who rated men on a stage by appearance only on first observation. It seems probable that degree of acquaintance or friendship after passing a certain threshold does

not affect the reliability of ratings. This is not the same as saying that observation does not increase the reliability of ratings. Cady finds, for instance, that observation is an important factor in increasing reliability.

When we investigate the reliability of self-estimate as compared with the ratings of others, we find that the results vary. Allport (3, p. 129) and H. L. Hollingworth (23) report that there is a larger average error in self-estimate than in the estimate of others. Cattell (16, p. 542) and Shen (95) report that the error in self-rating is less than in the rating of associates. Such divergent findings indicate that the difference is small, whichever way it lies.

Differences in ability to judge. Slawson (96) makes the point that two ratings by the same judges are no better than one. This is an important finding which should be verified, for, if true, it precludes trying to increase the reliability of ratings by having a judge repeat his own ratings. To increase the reliability of ratings, additional ratings must be made by independent observers.

Boyce (11), Cattell (17, p. 317), Hollingworth (51), Hughes (54), Lindsay (73), and others have noted the fact that individuals differ in their ability to rate accurately. For instance, if several persons have rated a group on some characteristic, and their ratings are averaged, this average may be taken as a tolerably accurate measure of the group. Of the several judges, some will correlate more highly with this average than others, indicating that some are better judges than others. In fact, it is commonly assumed that the individual whose judgment is most in harmony with the judgment of others is the most competent judge. Is this difference in ability to rate due to a general judicial capacity, to general intelligence, or to factors which are more specifically related to the particular situation in question? Hollingworth (50, 51) could detect the operation of no general judicial capacity. That is, an individual who is a good judge in one situation as compared with others may be a poor judge in another situation. The judge who correlates highest in his ratings with the "average" ratings on one trait may correlate comparatively low with the average rating on some other trait. Snow (97) tells us that business men are not necessarily better judges of the qualities of their employees than are other men.

Wells (121), as a result of his studies of literary merit, makes some valuable remarks about this general judicial capacity. He says that one is interested in ability to judge in a particular field rather than in the abstract power of judgment. To speak of judicial capacity is to refer to a faculty which psychology long ago discarded. Wells declares that one must have information as well as the ability to weigh it. That is, what counts is familiarity and expertness in the field in which we give our judgments rather than a general capacity to make judgments. Since part of the act of judgment is the noting of fine points and the making of distinctions, this can be achieved only through experience and interest in a particular field. Various findings by other investigators tend to throw light on this matter. Remmers and Place (88), and Cady (15) found that teachers agree better in rating students than students agree among themselves. It is the teacher's business to know his pupils and to be able to form estimates of them. In making these ratings teachers were doing something they were more skilled at than the student, even though the students knew each other more familiarly.

Arlett and Dowd (6) suggest that even trained judges may vary because traits function in different degrees in different situations. The class-room teacher, the physical education director, and the laboratory assistant have contacts with pupils in different situations and quite justly may form different opinions of the pupils on various traits. Webb (120) suggests that part of the unreliability of ratings is due to differences in points of view. The matter of definition of the traits to be rated, discussed in an earlier section, has a definite bearing on reliability of ratings. Again, trained judges may vary in their ratings because of personal likes and dislikes, and while this variation operates primarily to injure the validity of ratings, it also lowers the reliability of the ratings. Landis (72) points out that superficial physical characteristics influence judgment of deeper character qualities, with a resulting lessening in the value of ratings.

What are the characteristics of the good judge of others? This practical question has stimulated much thought. Those who do well in mental tests are not necessarily better judges of human nature within the limits of variability of the group studied. There are, however, two exceptions to this rule, as found by Hollingworth (23) and his students. Those who do well on mental tests

are good judges of intelligence and humor. As a generalization, we may affirm that, for certain admirable traits, there is a correlation between possession of a trait and ability to judge it. The more admirable a trait, the closer the relation between possession of it and ability to judge it. Also, one who knows himself best also knows others best in such desirable traits as neatness, intelligence, humor, refinement, and sociability. The reverse of this seems to be true in the case of undesirable traits. Those who possess a high degree of vulgarity, snobbishness, or conceit are disqualified to judge others on these traits. To rate desirable traits of character successfully, however, one should possess those traits himself to a high degree.

Adams (1) after studying extensively the characteristics of the good judge of personality, was led to differentiate the characteristics of the successful judge of self from the successful judge of others. He found the accurate judge of self somewhat more intelligent and more observing than the good judge of others. The good self-rater tends to be happier, less gloomy, less irritable, less liable to lose his head, more sympathetic, more generous, and more courageous than the good judge of others. But the outstanding characteristic of the good self-judge is his greater social interest and adaptability. Adams explains the apparent paradox, that the one who is most interested in others understands himself best and vice versa, as due to the fact that the person interested in others is best able to judge his own characteristics most impartially.

Hollingworth (50, p. 141) in his extensive investigations has brought out other interesting relationships. A group has been found to agree most with respect to their likes, while differing as individuals with respect to antipathies. For instance, a group which agrees in its general liking for certain types of jokes does not agree to the same extent in its aversion to types of jokes. On the other hand, though an individual shows offense or irritation at certain types of jokes rather uniformly, he does not respond so uniformly in the type of joke that does appeal to him. It is questionable whether this generalization will apply in other situations. Again, in a study of the persuasiveness of advertisements, Hollingworth (51) found that men agree most in their preferences, women in their dislikes. Is this a general sex difference? In another study Hollingworth reports that negative judgments are

more variable than positive judgments. These issues and Hollingworth's conclusions need further study and verification. There are many questions bearing on ability to render accurate judgments the solution of which would throw considerable light on the reliability of ratings.

The *consistency* of ratings, in the sense of self-consistency, is another phase of ability to rate. By *consistency* in this sense is meant the correlation between a judge's ratings at one time and those made at another time. Now being a consistent judge—agreeing with oneself on two different occasions—is not the same thing as agreeing with other people or with the true situation. Naturally one would expect an individual to agree more closely with his own judgment than with the judgment of another person or with the average judgment of many people. Hollingworth, in studying the judgments of persuasiveness of advertisements, found this to be true. The average correlation of the first and second judgment of an individual was .72, whereas the average correlation of judgments with the group average judgment was only .48.

Another related point is whether the consistency of rating for an individual is smaller or larger than the variability of judgments of a group. For instance, it is conceivable that in rating pictures or music for merit, a judge might be consistent in his own likes and dislikes but might vary widely from the likes and dislikes of others. On the other hand, it is conceivable that ten individuals might agree more closely in rating a group of applicants for a position than any one judge would agree in his own ratings on two different occasions. Wells (122, p. 547) records the following conclusions from his own study:

"We have thus made a study of variability in three classes of judgment, first, the highly subjective feeling of preference for different sorts of pictures, second, the more objective judgment of color differences, and finally a type of judgment whose accuracy could be readily measured by objective means. It appeared that in the first class the judgments of each individual cluster about a mean which is true for that individual only, and which varies from that of any other individual more than twice as much as its own judgments vary from it; that in the second class, with the colors, the variability of the successive judgments and those by different individuals markedly approached each other, but still preserved a significant difference; while in the third class, with

the weights, we found that there might be an excess of the individual variability over the 'social.' This comparison seems to afford, to a certain extent, a *quantitative criterion of the subjective.*"

Can judicial capacity be determined from consistency in rating? Is the person who on a second occasion agrees with his first judgment necessarily a good rater? Both Hollingworth (47, 51) and Slawson (96) answer no, that there is no substantial relation between capacity to rate and consistency of rating. The consistent judge may really be a poor judge. As Slawson says, "This relation was investigated with the intention of ascertaining whether it was possible to determine capacity from consistency, since the latter can be determined so quickly and easily. The results were negative."

It has been found by Hollingworth and Cady (15) that reliability varies with the degree of confidence in the ratings. If a judge in making his ratings is asked to indicate opposite each rating the degree of certainty with which he makes the rating, the ratings recorded with confidence have been found to be the more reliable. High reliability of ratings may be obtained by selecting only those ratings of which the judge feels sure,—a fact of importance in using ratings in experimental work.

Related to this last point is the discovery made by Cady (15) that *certain* ratings are apt to be *extreme* ratings. We are usually more sure of our judgment when we rate a person high or low than when we give him an intermediate position. Indeed, intermediate ratings are frequently given simply because the rater knows little one way or the other about the person.

Finally it has been verified by Cattell, Hollingworth (47), Snow (97), Cady (15), Cox (26), Wells (122), and others that extreme ratings are more reliable than intermediate ratings. Cox (26) found, for instance, that the higher the IQ of a person, the more reliably that IQ could be estimated. Cattell's (16) study of eminent scientists revealed that there was much more agreement as to the rank of the first few scientists than for scientists who came far down the list. It would seem that in proportion as a person becomes outstanding in any trait or quality, so much the finer are the discriminations that are made concerning him, probably because differences in ability or quality between individuals are greater at the extremes. Genius is usually subjected to ruth-

104 Diagnosing Personality and Conduct

less scrutiny, simply because it is outstanding, whereas the average individual passes by unnoticed. It was discovered both by Hollingworth (47 ch. IV) and in the California study of gifted children (103) that there is regression in the evaluation of extreme cases, the tendency being to underrate the superior individuals and to overrate the inferior individuals.

Differences in the traits being rated. First it should be noted that Yoakum and Manson (123) report that ratings on synonymous traits correlate well.

Several investigators have established the point that it is more difficult to rate certain traits than others. Slawson (96), for instance, finds for ten traits the following average reliability coefficients.

TABLE 15
RELIABILITY COEFFICIENTS OF RATINGS FOR TEN TRAITS
(from Slawson)

All-round value to service603
Coöperativeness522
Leadership503
Effort491
Understanding of children472
Professional interest and growth470
Appearance460
Tact453
Punctuality408
Judicial sense335

Cattell (16) found the relative agreement of judgments when twelve scientific men estimated the character of five of their colleagues. Norsworthy (81), using the same list of traits, had nine members of a college sorority judged by five of their intimate acquaintances. Hollingworth (49, p. 79) has reduced their results to comparative figures which represent the amount of disagreement among judges in estimating traits of character, the average disagreement being taken as 100 (see Table 16).

Hollingworth (23, p. 173) in one of his investigations obtained data which shed light on the relative reliability with which traits are rated. In a class experiment twenty-five students rated themselves and each other. Table 17 on the following page, for which ranking was the method used, shows the average deviation of judgment of twenty-four acquaintances on nine traits. The average deviation for a purely chance arrangement of items is 6.

TABLE 16

AMOUNT OF DISAGREEMENT AMONG JUDGES IN ESTIMATING, ON THE BASIS OF
ACQUAINTANCE, THE TRAITS OF OTHERS, IN TWO INVESTIGATIONS

(from H. L. Hollingworth, *Judging Human Character*, p. 79)

Trait	RELATIVE DIVERGENCE OF JUDGES			Classifi- cation
	Cattell (12 judges)	Norsworthy (5 judges)	Average of both	
Efficiency	75	92	83	Class A Median 89
Originality	95	77	86	
Perseverance	75	101	88	
Quickness	90	88	89	
Judgment	100	78	89	
Clearness	104	75	90	Close Agreement
Energy	75	109	91	
Will	85	98	91	
Mental balance	110	81	96	Class B Median 100
Breadth	100	92	96	
Leadership	90	103	96	
Intensity	85	113	99	
Reasonableness	115	86	100	
Independence	104	98	101	Fair Agreement
Refinement	90	116	103	
Physical health	115	92	103	
Emotions	120	91	105	
Courage	100	119	109	Class C Median 118
Unselfishness	115	106	110	
Integrity	104	130	117	
Coöperativeness	125	113	119	
Cheerfulness	130	112	121	
Kindliness	120	125	123	Poor Agreement

TABLE 17

AVERAGE DEVIATION OF TRAITS IN RANKING
(from Hollingworth)

	Average deviation in rank
Neatness	4.5 steps
Intelligence	3.7
Humor	4.5
Conceit	4.1
Beauty	3.8
Vulgarity	3.5
Snobbishness	4.8
Refinement	5.9
Sociability	4.7
Average	4.4 steps

In Miner's study (78, p. 127), previously referred to, reliability coefficients are given for several traits. The figures quoted here are for two judgments against two others, rating from thirty to thirty-six individuals.

TABLE 18
RELIABILITY COEFFICIENTS OF RATINGS FOR SIX TRAITS
(from Miner)

	<i>Reliability coefficients</i>
Energy78
Leadership77
General ability76
Reliability71
Initiative67
Common sense62

Shen (94) in his study presents reliability coefficients for thirteen judges' ratings on eight traits for twenty-eight individuals. As averaged by the writer, they are as follows:

TABLE 19
RELIABILITY COEFFICIENTS OF RATINGS FOR EIGHT TRAITS
(from Shen)

	<i>Reliability coefficients</i>
Scholarship71
Leadership68
Intellectual quickness68
Intellectual profoundness64
Memory55
Persistence40
Adaptability38
Impulsiveness34

Enough evidence has been given to show that certain traits may be rated more reliably than others. The reliability with which a trait may be rated is dependent in large part on the objectivity of the trait, and vice versa, the objectivity of a trait may be estimated by the degree of reliability with which it can be rated. The more objective the trait, the less two people will disagree in making judgments on it. In general, traits which somehow leave their mark on things or influence external events are more reliably rated than qualities which are merely characteristics of the person being judged. Hollingworth (49, p. 80) differentiates be-

tween the two extremes as follows: "The A traits we may designate as 'objective' in the sense that they represent reactions to objects and impersonal situations and tasks, and are likely to result in objective products such as inventions, factories, books, bank accounts, salaries, positions, records, etc. These objective products are definite manifestations of the traits in question, and they are open to general inspection. The C traits, on the other hand, represent reactions to the presence and character of other persons. They are personal, social, moral, they do not so definitely produce objective products open to general inspection. Instead, they lead mainly to personal and emotional reactions on the part of others; hence we may designate them 'subjective' traits." Cattell has noted that it is easiest to rate people's reactions to objective things and hardest to rate reactions to other people. Slawson (96) in his work found that the "uniformity, explanations, and accessibility of criteria" for defining a trait were the most important factors in determining the objectivity of the trait. Since different traits differ so markedly in the degree to which they may be reliably judged, it behooves one who is planning a rating card to consider the items to be included from this standpoint. Such items as coöperativeness, cheerfulness, kindliness, refinement, adaptability, and impulsiveness should be avoided in favor of more objective and better-defined traits.

Hollingsworth (47, p. 112) discovered that the longer the series, the less accurately any item in the series was placed in rank by a series of judges. Confusion is apt to result if the number of individuals to be ranked becomes too large. It is never well to require persons to rank more than a dozen or fifteen persons on any trait.

Rugg (90) and also Slawson (96) have pointed out that a general all-round trait is more reliably rated than a more specific trait. This is probably caused by the so-called halo effect. Ratings on a narrow trait require more detailed observation than is usually made. At any rate, as was mentioned on page 95, the combination of ratings of specific traits that correlate highly with a general trait is considerably more reliable than the blanket rating of a general trait. In this connection Hanna shows that the correlation between ratings of teachers of the same subject is higher than the ratings of teachers of different subjects for the same trait. That is, as the observation of situations becomes narrow and

more uniform, judges will agree in their judgments. Cady (15), for instance, finds that observation increases the reliability of ratings. The long and thorough observation which should precede ratings, discussed earlier in this chapter, is one of the surest means of increasing rating reliability.

Validity of Ratings

In a certain sense there is nothing more valid than a judgment. In the final analysis, all of our knowledge has its origin in observation and in interpretations made of observations. However, we have just seen that human judgment is liable to error, and that variable or compensatory errors affect the accuracy of the rating without impugning its genuineness or honesty. In the present section we are to consider the question as to whether ratings really are measures of what they set out to measure. We shall find that ratings are also subject to certain constant errors and that apart from the inaccuracies called accidental or chance, there are other serious errors that we may call systematic. We shall find that human judgment is not strictly impartial, but has a tendency to err through partiality and favoritism.

The acquaintance factor. It is strange that acquaintance should sometimes be a factor in decreasing the validity of ratings. Throughout this chapter the need for observation and more adequate information has been stressed. There is perhaps no factor more important in helping to reduce the variable errors of ratings than familiarity. Hence we may say in one sense that acquaintance makes a judge more competent. We all feel that a teacher is a better judge of the members of her class than of any other class in the school and that every parent is a better judge of his own children than of his neighbor's children. But with acquaintance there creeps in an insidious tendency to become lenient or to show favoritism, or at least to be off one's guard. Knight (64) first emphasized the systematic error that results from long acquaintance. He found that there was a distinct tendency for supervisors to rate the teachers highest whom they had known longest, where impartial testimony would not credit these older teachers with greater teaching ability. It would seem that with long acquaintance there goes an unconscious tendency to excuse, to explain away defects or to overlook deficiencies that

would be considered in an individual less well known. Shen (93) finds that the same tendency to overrate acquaintances holds among college classmates. This subtle factor, which works unconsciously, must be recognized and allowed for in using ratings in personnel work. In the rating of teachers by supervisors or the rating of employees by managers this factor is certain to slip in to decrease the value of ratings.

Self-rating. Another factor which influences the validity of ratings is brought to light in studies of self-rating. Cattell (16, p. 542) avers that there is no constant error in judging ourselves; and, if we consider all kinds of persons and all kinds of traits, we may agree that in a certain sense this is so. Although later investigators have noticed a definite tendency to overrate oneself, it can be pointed out that their studies dealt almost entirely with desirable traits. Finally Hollingworth (23) has discovered that there is a tendency to *overrate* the self on desirable traits and to *underrate* the self on *undesirable* traits. Hurlock (56) found that of 12,690 responses made by 423 children in selecting traits to describe themselves, only 6 per cent were ones relating to socially undesirable traits. The error in rating was always in favor of the person doing the rating as compared with what was obtained in his estimates of other persons.

We do not begin to know all we would like to know about self-rating. Many factors influence the results, and several of the studies are contradictory. Shen (95) finds that intellectual *profoundness* and memory are underestimated to a greater extent than impulsiveness, although the former traits would hardly be called less desirable than impulsiveness. Shen suggests that a defense mechanism is at play here, and that instead of overrating there is a tendency to excuse the self. *Self-rating* may therefore be a key to the diagnosis of types of maladjustment which are familiar both in schools and in industry. Shen also finds that this tendency to overestimate or underestimate the self is more or less consistent with the individual, depending more upon the individual than upon the trait. This is added evidence that aberrations of self-estimate may be symptomatic of maladjustments of the personality.

Trow and Pu (114) believe that the tendency to overrate the self on desirable traits is a national characteristic. They find that thirteen to sixteen out of eighteen Chinese college students whom

they studied tended to underrate themselves on desirable traits.

A single experiment by Hollingworth (52) brings out two other phases of self-rating which he calls the factors of *optimism* and *altruism*. A number of persons were set to work by Hollingworth at the continuous performance of a series of mental and physical tests. After each trial the performer was required to judge whether he had done better or worse than usual on this occasion. In each case another person was required to watch the performer, and to judge, in the capacity of witness, whether the performance had been better or worse than usual for the individual who was doing the work. Hollingworth found that a person has a tendency to judge himself as doing better than usual, thereby exhibiting a kind of optimism. Witnesses, however, showed an even greater inclination to judge a performance as better than typical, thus evidencing not only optimism but altruism.

This tendency to favor the self spreads to one's friends or to members of one's group. Kinder (58), for instance, finds that there is a tendency to overrate members of the same sex as compared with members of the other sex. Hart and Olander (38) tell us that men are more lenient in their ratings than are women. Remmers and Place (88) find that students rate each other higher on the average than teachers rate them. Cattell (16, p. 542) finds in his study of eminent scientists that a professor is apt to overestimate the standing of his colleagues. There is a tendency for parents to overestimate the attainments of their own children. There seems to be a general tendency to overrate not only ourselves on desirable traits but also those with whom we are associated or in whom we have an interest.

So important does this matter of self-rating loom that Yoakum and Manson believe that the relative desirability of traits may be determined from amount of overestimation in self-ratings.

Another phase of self-rating was brought out by Hoffman (46), who discovered that a person possessing a desirable trait to high degree usually underestimates his possession of it, whereas a person deficient in a desirable trait overestimates his possession of it to an even greater extent. For instance, a pupil standing high in his class in school will modestly place himself somewhat lower in standing, whereas a low-standing pupil will rate himself higher. Allport (3) in corroboration tells us that the intelligent

underrate themselves less than the unintelligent overrate themselves. Even though parents tend to overestimate their children's attainments, parents of gifted children usually underestimate their children's gifts, Terman (103) finds, since they place a lower estimate on the children's abilities than do their teachers. Hoffman also finds that persons thought conceited by others are the more apt to overestimate themselves on desirable traits.

Halo effect. Many investigators have noted the tendency for general impressions to spread to specific traits, to which tendency Thorndike (107) has given the name "halo effect." Wells, in his *Statistical Study of Literary Merit* (121, p. 21), was the first to mention this phenomenon. He says, "There is a possibility of one rather disturbing constant error in measures of this nature, whose extent it is never possible to know accurately. There is noted introspectively a tendency to grade for general merit at the same time as for the qualities, and to allow an individual's general position to influence his position in the qualities. This would be the case especially in the case of those qualities that were ill-defined in the minds of the subjects, and tended to be interpreted rather in terms of general merit. We might thus have a grading of charm by general merit instead of general merit by charm." Webb (120) notes the same influence: "Let us suppose, for instance," he says, "that the observers, in estimating the intelligence qualities, are biased in the direction of marking subjects who possess other desirable qualities too highly and vice versa."

The halo effect was noted to be a prominent factor in connection with the work with the Army Rating Scale. Thorndike (107), the first to interpret these findings during the World War, says, "The magnitude of the constant error of the halo, as we have called it, also seems surprisingly large, though we lack objective criteria by which to determine its exact size."

Rugg has described in detail the findings in connection with the Army Rating Scale. He says (90, p. 37), "We judge our fellows in terms of a general mental attitude toward them; and there is dominating this mental attitude toward the personality as a whole, a like mental attitude toward particular qualities." Rugg tells of a certain "Captain X" who was so well known and conspicuous in his group that he was used by thirteen officers in twenty different subordinate scales—physical qualities, intelligence, leadership, etc.—as "scale" man on the Army Rating Scale as

"the poorest man I ever knew." Yet this same Captain X stood first on three different psychological tests among 151 officers. He had been a Rhodes scholar from a Middle Western State university, and at Oxford he had made such a record that he was excused from certain examinations. Comments of eight of the thirteen officers who had judged him so severely showed that their estimates of his intelligence, his physical qualities, and his leadership were dominated by their opinions of his personal qualities. They were unanimous in saying that it was impossible "to live with him." He was a "rotter," or "yellow," or a "knocker," or "conceited." "The suggestion comes insistently, however, that one of the most potent influences working against accurate estimates of character is the prevalence of just such general attitudes toward our associates and subordinates."

Knight (65) also has found that the halo of general impression is a considerable factor working to lower the validity of ratings. In his work in the field of estimating teaching success, Knight showed that suspiciously high correlations prevailed between the ratings of different traits. The same phenomenon was disclosed in an earlier study by Boyce (11). Knight concludes, "This decided monotony of the size of the correlations, which are obviously too high, is potent witness of the presence of spread of general estimate. . . . Their monotonous similarity also suggests that, when analyzed judgments are attempted, the influence of general estimate is so strong that the resulting analyses are perhaps even more justifications of the general estimate than they are judgments of the specific trait."

Hollingworth (23) warns that this spread of general estimate should not be confused with general "stand-outishness." Certain individuals do stand out prominently in many respects. Correlation is the general law of relationship between desirable traits, so that an individual possessing a desirable characteristic will be found to possess other desirable characteristics also.

The present writer (101) attempted to estimate the size of the halo effect by having two teachers rate a group of forty pupils on seven traits. A composite rating for each child, found by adding together the ratings of the seven traits by one rater, was taken as the general impression rating. Partial correlations were computed, holding constant the general impression of each of the raters. The resulting partial correlation between any two traits

was, as it were, freed from the halo effect or the influence of general impression. The average coefficients of reliability as found were .390 and .377. The average partial coefficients of reliability were .145 and .200. The differences, .245 and .177, serve as a measure of the influences of the general impression. The study assigns five possible reasons for a large halo effect in the rating of any trait or habit: (1) the trait or habit is one which is not *easily* observed; (2) the trait or habit is one which is not commonly observed or thought about, such as one which is not usually emphasized in the classroom; (3) the trait or habit is not clearly defined; (4) the trait or habit is one which involves reactions with other people rather than "self-contained" behavior; (5) the trait or habit is one with high moral importance in its usual connotation. Knight believes that acquaintance increases the halo effect.

Hartshorne and May report extensively on the correlation between their performance tests and ratings. They found (39) that measures of class-room honesty correlated with a rating for general honesty around .40. With improved rating techniques they found (40) a correlation of .61 (corrected for attenuation) between total score for service on performance tests and total reputation for service by ratings. The correlation of the reputations with separate service tests was much lower. The corresponding correlation between tests and ratings of self-control was .52. They demonstrate (41) that although the average correlation between actual judgments and objective tests is only .35, it is possible by taking the sum of ratings obtained in different ways and the sum of tests on various characteristics to obtain much higher correlations—i.e., between .50 and .60. With a wider variety of tests, several sets of ratings, and the assumption of perfectly accurate measurement, they demonstrate that the correlation would approach unity. In short, though actual ratings agree only indifferently with tests, it is theoretically possible by extending both ratings and testing to cover a wider variety of situations until the agreement would become as close as the accuracy of judgment and measurement permits.

Conclusion

Enough is now known about ratings so that skilful and proper use may be made of this method of measurement. Time was when

114 Diagnosing Personality and Conduct

ratings were indiscriminately used with no thought of safeguards or precautionary techniques. Unfortunately this holds true to-day among novices. Directly after the World War a reaction set in, due to experience with the Army Rating Scale and also particularly to Rugg's admissions concerning his experiences with rating as set forth in his series of articles entitled "Is the Rating of Human Character Practicable?" But experimentation continued, and refinements were made until to-day ratings are accepted, though somewhat gingerly, as a valid means of obtaining data. But the war experience taught us to be alert against the unreliability of the single offhand rating.

Ratings are capable of great improvement if the following factors, among others, are considered and if enough pains are taken to ensure that the ratings are reliable:

1. Ratings should be made in a systematic way.
2. An extended period of observation should precede rating.
3. It should be kept in mind that rating is something in which the rater may improve through practice just as he grows more skilful in judging the quality of handwriting or an English composition through practice.
4. More attention should be paid to defining the qualities or traits to be rated, and more extensive definitions should be introduced. In place of the man-to-man rating scheme, a definite and extended description of the scale men might prove effective (a suggestion borrowed in part from experience with the English composition scale technique).
5. Single ratings should not be used in the rating of human qualities. Sufficient reliability may be obtained only when a composite is made of the independent judgment of from five to ten observers.
6. For experimental purposes all ratings should be discarded except those which are at the extreme ends of the rating scale and those on which the raters are sure of their judgment.
7. Traits for rating should be selected which experience shows yield better than average reliability.
8. So far as possible bias should be eliminated from ratings. Individuals should not be expected to give fair ratings when judging themselves, friends, old acquaintances, or persons whom they much like or dislike, admire or despise.

REFERENCES

1. ADAMS, H. F., "The Good Judge of Personality," *Journal of Abnormal and Social Psychology*, 22: 172-181 (1927).
2. ATHEARN, W. S., and others, "Measurements and Standards in Religious Education," *The Indiana Survey of Religious Education* (1924), 532 pp.
3. ALLPORT, F. H., *Social Psychology* (Houghton Mifflin Company, 1924), 453 pp.
4. ALLPORT, G. W., "Personality and Character," *Psychological Bulletin*, 18: 441-455 (1921).
5. ARLETT, A. H., and DOWD, C. E., "Variability Among a Group of Judges in Rating Character Traits in Children," *Psychological Bulletin*, 23: 617-19 (Nov., 1926).
6. BARR, A. S., "A Study in Social Rating," *School and Home Education*, 41: 46-48 (Oct. and Nov., 1921).
7. BARRETT, M., "The Order of Merit Method and the Method of Paired Comparisons," *Journal of Philosophy*, 10: 382-384 (July 3, 1913). See also *Psychological Review*, 21: 278-294 (1914).
8. BENGE, E. J., "Are Rating Scales Justified?" *Industrial Management*, 65: 199-202 (1923).
9. BILLS, M. A., "A Method for Classifying the Jobs and Rating the Efficiency of Clerical Workers," *Journal of Personnel Research*, 1: 384-393 (1923).
10. BINGHAM, W. V., and FREYD, M., *Procedures in Employment Psychology* (A. W. Shaw Company, 1926).
11. BOYCE, A. C., "Methods for Measuring Teachers' Efficiency," *14th Yearbook, National Society for the Study of Education* (1915), Part II.
12. BRADSHAW, F. F., "The American Council on Education Rating Scale," *Archives of Psychology*, 18, No. 119 (1930).
13. BROWN, W. S., "A Note on the Psychogalvanic Reflex Considered in Conjunction with Estimates of Character Qualities," *British Journal of Psychology* (General Section), 16, Part 2: 130-135 (1925).
14. BURTT, H. E., *Employment Psychology* (Houghton Mifflin Company, 1926).
15. CADY, V. M., *The Estimation of Juvenile Incurability*, *Journal of Delinquency Monographs*, No. 2 (April, 1923).
16. CATTELL, J. McK., *American Men of Science* (2d edition, 1910), Appendix.
17. ———, "Statistics of American Psychologists," *American Journal of Psychology*, 14: 310-328 (1903).
18. *Character Education Methods—The Iowa Plan* (\$20,000 Award, 1922) (Washington, D. C., Character Institution, 1922), 46 pp.

19. CHASSELL, C. F., UPTON, S. M., and CHASSELL, L. M., "Scales for Measuring Habits of Good Citizenship," *Teachers College Record*, 23: 52-79 (Jan. 1922).
20. CHU, J. P., *Chinese Students in America; Qualities Associated with Their Success*, Teachers College Contributions to Education, No. 127 (1922).
21. CLEETON, G. U., "Originality: A Summary of the Experimental Literature," *Journal of Abnormal and Social Psychology*, 21: 304-315 (Oct., Dec., 1926).
22. CLEETON, G. U., and KNIGHT, F. B., "Validity of Character Judgments Based on External Criteria," *Journal of Applied Psychology*, 8: 215-231 (June, 1924).
23. COGAN, L. C., CONKLIN, A. M., and HOLLINGWORTH, H. L., "An Experimental Study of Self Analyses, Estimates of Associates and Results of Tests," *School and Society*, 2: 171-179 (1915).
24. CONKLIN, E. S., and SUTHERLAND, J. W., "A Comparison of the Scale of Values Method with the Order-of-Merit Method," *Journal of Experimental Psychology*, 6: 44-57 (1923).
25. CONNOR, W. L., "A New Method of Rating Teachers," *Journal of Educational Research*, 1: 338-358 (1920).
26. COX, K., *Genetic Studies of Genius*, II, *The Early Mental Traits of Three Hundred Geniuses* (Stanford University Press, 1926).
27. DORCUS, R. M., "Some Factors Involved in Judging Personal Characteristics," *Journal of Applied Psychology*, 10: 502-518 (Dec., 1926).
28. DOWNEY, J., *Study of Family Resemblances in Handwriting*, Bulletin No. 1, Department of Psychology (University of Wyoming, 1910).
29. EARLE, M. G., "The Relation between Personality and Character Traits and Intelligence," *Journal of Applied Psychology*, 10: 453-461 (Dec., 1926).
30. FERNALD, M. R., and others, *A Study of Women Delinquents in New York State* (The Century Co., 1920), 542 pp.
31. FILER, H. A., and O'Rourke, L. J., "Progress in Civil Service Tests," *Journal of Personal Research*, 1: 484-520 (March, 1923).
32. FREYD, M., "The Graphic Rating Scale," *Journal of Educational Psychology*, 14: 83-102 (Feb., 1923).
33. ———, "Graphic Rating Scale for Teachers," *Journal of Educational Research*, 8: 433-439 (1923).
34. FURFEY, P. H., "An Improved Rating Scale Technique," *Journal of Educational Psychology*, 17: 45-48 (Jan., 1926).
35. GALLUP, G. H., "Traits of Successful Retail People—A Pre-

- liminary Study," *Journal of Personal Research*, 4: 474-482 (April, 1926).
36. GARRETT, H. E., "An Empirical Study of Various Methods of Combining Incomplete Order of Merit Ratings," *Journal of Educational Psychology*, 15: 157-172 (1924).
 37. HANNA, J. V., "Variable Factors Encountered in the Rating of Students," *School Science and Mathematics*, 25: 481-488 (1925).
 38. HART, H. N., and OLANDER, E., "Sex Differences in Character as Indicated by Teachers' Ratings," *School and Society*, 20: 381-382 (1924).
 39. HARTSHORNE, H., and MAY, M. A., *Studies in Deceit* (The Macmillan Company, 1928).
 40. HARTSHORNE, H., MAY, M. A., and MALLER, J. B., *Studies in Service and Self-Control* (The Macmillan Company, 1929).
 41. HARTSHORNE, H., MAY, M. A., and SHUTTLEWORTH, F. K., *Studies in the Organization of Character* (The Macmillan Company, 1930).
 42. HAUGHT, B. A., "A Scheme for Combining Incomplete Rankings," *Journal of Applied Psychology*, 7: 168-172 (1923).
 43. HAYES, M. S., and PATERSON, D. G., "Experimental Development of the Graphic Rating Method," *Psychological Bulletin*, 18: 98-99 (1921).
 44. HEPNER, T. W., "Better Judgments of Men," *Industrial Psychology*, 1: 19-24 (January, 1926).
 45. HILLEGAS, M. B., "A Scale for the Measurement of Ability in English Composition," *Teachers College Record*, 13, No. 4: 1-54 (Sept., 1912).
 46. HOFFMAN, G. J., "An Experiment in Self-Estimation," *Journal of Abnormal and Social Psychology*, 18: 43-49 (April, June, 1920).
 47. HOLLINGWORTH, H. L., *Experimental Studies in Judgment*, Archives of Psychology, No. 29 (Dec., 1913), p. 116.
 48. ———, "Influence of Form and Category on the Outcome of Judgment," *Journal of Philosophy*, 9: 513-520 (1912).
 49. ———, *Judging Human Character* (D. Appleton and Company, 1922).
 50. ———, "Judgments of the Comic," *Psychological Review*, 18: 132-156 (1911).
 51. ———, "Judgments of Persuasiveness," *Psychological Review*, 18: 234-256 (1911).
 52. ———, *Vocational Psychology* (D. Appleton and Company, 1917), 308 pp.
 53. HUGHES, W. H., "A Rating Scale of Individual Capacities, Attitudes and Interests," *Journal of Educational Method*, 3: 56-65 (Oct., 1923).

54. HUGHES, W. H., "General Principles and Results of Rating Trait Characteristics," *Journal of Educational Method*, 4: 421-431 (1925).
55. HUGHES, W. H., "Relation of Intelligence to Trait Characteristics," *Journal of Educational Psychology*, 17: 482-494 (Oct., 1926).
56. HURLOCK, E. B., "A Study of Self-Ratings by Children," *Journal of Applied Psychology*, 11: 490-502 (1927).
57. KELLY, R. W., *Training Industrial Workers* (The Ronald Press Company, 1920), 437 pp.
58. KINDER, V. S., "Through Our Own Looking Glass," *School and Society*, 22: 533-534 (1925).
59. KINGSBURY, F. A., "Analyzing Ratings and Training Raters," *Journal of Personal Research*, 1: 377-383 (Dec., Jan., 1922-1923).
60. ———, "Making Rating Scales Work," *Journal of Personnel Research*, 4: 1-6 (May, 1925).
61. ———, "The Principles Involved in Securing Service Ratings as Exemplified in a Large Bank," *Pub. Personnel Studies*, 3: 70-84 (1925).
62. KITSON, H. D., *The Psychology of Vocational Adjustment* (J. B. Lippincott Company, 1925), 273 pp.
63. KNIGHT, F. B., "Analysis of Teaching and Teachers," *Journal of Educational Research*, 10: 222-227 (1924).
64. ———, "The Effect of the 'Acquaintance Factor' upon Personal Judgments," *Journal of Educational Psychology*, 14: 129-142 (March, 1923).
65. ———, *Qualities Related to Success in Teaching*, Teachers College Contributions to Education, No. 120 (1922), 67 pp.
66. KNIGHT, F. B., and FRANZEN, R., "Pitfalls in Rating Schemes," *Journal of Educational Psychology*, 13: 204-213 (1922).
67. KORNHAUSER, A. W., "A Comparison of Raters," *Journal of Personnel Research*, 5: 338-344 (Jan., 1927).
68. ———, "A Comparison of Rating Different Traits," *Journal of Personnel Research*, 5: 440-446 (March, 1927).
69. ———, "A Statistical Study of a Group of Specialized Office Workers," *Journal of Personnel Research*, 2: 103-123 (1923).
70. ———, "Reliability of Average Ratings," *Journal of Personnel Research*, 5: 309-317 (Dec., 1926).
71. ———, "What Are Rating Scales Good For?" *Journal of Personnel Research*, 5: 189-193 (1926).
72. LANDIS, C., "The Justification of Judgments. A Study of the Reasons Given by Raters in Support of Their Judgments of

- Emotionality, Stability and Expressiveness," *Journal of Personnel Research*, 4: 7-19 (May, 1925).
73. LINDSAY, E. E., "Personal Judgments," *Journal of Educational Psychology*, 12: 413-415 (1921).
 74. MAGSON, H., *How We Judge Intelligence*, British Journal of Psychology, Monograph Supplements, IX (1926).
 75. MANN, C. R., *A Study of Engineering Education*, Carnegie Foundation for Advancement of Teaching, Bulletin No. II (1918), 139 pp.
 76. MARSHALL, L. C., *Business Administration* (University of Chicago Press, 1921).
 77. MEIER, N. C., "A Study of the Downey Test by the Method of Estimates," *Journal of Education Psychology*, 14: 385 (Oct., 1923).
 78. MINER, J. B., "Evaluation of a Method for Finely Graded Estimates of Abilities," *Journal of Applied Psychology*, 1: 123-133 (June, 1917).
 79. MOORE, B. V., *Personnel Selection of Graduate Engineers*, Psychological Monographs, No. 30 (1921), 84 pp.
 80. MOORE, H. T., "Three Types of Psychological Rating in Use with Freshmen at Dartmouth," *School and Society*, 13: 418-420 (1921).
 81. NORSWORTHY, N., "The Validity of Judgments of Character," *Essays in Honor of William Jones* (2d edition, 1910), pp. 552-567.
 82. PATERSON, D. G., "Methods of Rating Human Qualities," *The Annals of the American Academy of Political and Social Science*, November, 1923, Vol. 110, No. 199, 81-93.
 83. ———, "The Scott Company Graphic Rating Scale," *Journal of Personnel Research*, 1: 361-376 (1923).
 84. PATERSON, D. G., and RUMI, B., "The Extension of Rating Theory and Scale Technique," *Psychological Bulletin*, 17: 80-81 (1920).
 85. PEARSON, K., "On the Relationship of Intelligence to Size and Shape of Head, and to Other Physical and Mental Characters," *Biometrika*, 5: 105-146 (1906-1907).
 86. POFFENBERGER, A. T., *Psychology in Advertising* (A. W. Shaw Company, 1925).
 87. REAM, M. J., "A Statistical Method for Incomplete Order of Merit Ratings," *Journal of Applied Psychology*, 5: 261-267 (1921).
 88. REMMERS, H. H., and PLACE, M. J., "Reliability of Ratings at Purdue University," *Industrial Psychology*, 1: 717 (Nov., 1926).
 89. REMMERS, H. H., SHOCK, N. W., and KELLY, E. L., "An Empirical Study of the Validity of the Spearman-Brown For-

- mula as Applied to the Purdue Rating Scale," *Journal of Educational Psychology*, 18: 187-195 (March, 1927).
90. RUGG, H. O., "Is the Rating of Human Character Practicable?" *Journal of Educational Psychology*, 12: 425-438, 485-501 (1921); 13: 30-42, 81-93 (1922).
 91. ———, "Self-Improvement of Teachers through Self-Rating," *Elementary School Journal*, 20: 670-684 (1920).
 92. SCOTT, W. D., and CLOTHIER, R. C., *Personnel Management* (A. W. Shaw Company, 1923).
 93. SHEN, E., "The Influence of Friendship upon Personal Ratings," *Journal of Applied Psychology*, 9: 66-68 (March, 1925).
 94. ———, "The Reliability Coefficient of Personal Ratings," *Journal of Educational Psychology*, 16: 232-236 (April, 1926).
 95. ———, "The Validity of Self-Estimate," *Journal of Educational Psychology*, 16: 104-107 (Feb., 1925).
 96. SLAWSON, J., "The Reliability of Judgments of Personal Traits," *Journal of Applied Psychology*, 6: 161-171 (1922).
 97. SNOW, A. J., "An Experiment in the Validity of Judging Human Ability," *Journal of Applied Psychology*, 8: 339-347 (1924).
 98. SPAULDING, E. R., "The Role of Personality Development in the Reconstruction of the Delinquent," *Journal of Abnormal and Social Psychology*, 16: 97-114 (June-Sept., 1921).
 99. STENQUIST, J. L., "An Improved Form of Rating by the Order of Merit Method," *Journal of Educational Psychology*, 1: 526-528 (Dec., 1920).
 100. SWEET, L., *The Measurement of Personal Attitudes in Younger Boys*, Young Men's Christian Association, Occasional Studies, No. 9 (1929).
 101. SYMONDS, P. M., "Notes on Rating," *Journal of Applied Psychology*, 7: 188-195 (1925).
 102. ———, "On the Loss of Reliability Due to Coarseness of the Scale," *Journal of Experimental Psychology*, 7: 456-460 (1924).
 103. TERMAN, L. M., *Genetic Studies of Genius* (Stanford University Press, 1925), Vol. 1, 648 pp.
 104. ———, "The Army Rating Scale," *Psychological Bulletin*, 15: 177-187 (1918).
 105. ———, *The Intelligence of School Children* (Houghton Mifflin Company, 1919), 313 pp.
 106. THOMPSON, G., "A Rating Scale for Teaching Ability in Students," *Journal Experimental Pedagogy*, 6: 75-82 (1921).
 107. THORNDIKE, E. L., "A Constant Error in Psychological Rating," *Journal of Applied Psychology*, 4: 25-29 (1920).

108. THORNDIKE, E. L., "Handwriting," *Teachers College Record*, 11, No. 2 (March, 1910), 93 pp.
109. ———, "The Technique of Combining Incomplete Judgments of the Relative Position of 'n' Facts Made by 'n' Judges," *Journal of Philosophy*, 13: 197-204 (1916).
110. THORNDIKE, E. L., BREGMAN, E. A., and COBB, M., "The Selection of Tasks of Equal Difficulty by a Consensus of Opinion," *Journal of Educational Research*, 9: 133-139 (1924).
111. THURSTONE, L. L., "An Experimental Study of Nationality Preferences," *Journal of General Psychology*, 1: 405-425 (1928).
112. ———, "The Method of Paired Comparisons for Social Values," *Journal of Abnormal and Social Psychology*, 21: 384-400 (Jan.-March, 1927).
113. TIEGS, E. W., *An Evaluation of Some Techniques of Teacher Selection* (The Public School Publishing Company, 1928).
114. TROW, W. C., and PU, A. S. T., "Self-Ratings and the Chinese," *School and Society*, 26: 213-216 (1927).
115. TYLER, H. T., *A Study of the Bearing of Certain Personality Factors Other Than Intelligence upon Academic Success in Teachers College*, unpublished doctor's dissertation, Columbia University, 1930.
116. WAGNER, C. A., "The Construction of a Teacher Rating Scale," *Elementary School Journal*, 21: 361-366 (1921).
117. WAITE, H., "The Teacher's Estimation of the General Intelligence of School Children," *Biometrika*, 8: 79-93 (July, 1911).
118. WATSON, G. B., *Some Methods of Program Appraisal*, Program Papers, No. 4 (Association Press, 1926), 32 pp.
119. ———, "The Measurement of the Less Tangible Qualities," *Vocational Guidance Magazine*, 4: 281-289 (1926).
120. WEBB, E., *Character and Intelligence*, British Journal of Psychology Monographs, No. 3 (1915), 99 pp.
121. WELLS, F. L., *A Statistical Study of Literary Merit*, Archives of Psychology, No. 7 (1907).
122. ———, "On the Variability of Individual Judgments," *Essays in Honor of William James* (1908), p. 511.
123. YOAKUM, C. S., and MANSON, G. E., "Self-Ratings as a Means of Determining Trait Relationships and Relative Desirability of Traits," *Journal of Abnormal and Social Psychology*, 21: 52-64 (1926).

Chapter IV

THE QUESTIONNAIRE

JUST as the *test* is the instrument which one uses to test *ability*, so the *questionnaire* is the instrument best fitted to measure *conduct*. Although tests and questionnaires are sometimes confused, the distinction between them should be sharply drawn. All research instruments are loosely called "tests" in recent experimental literature. Although both tests and questionnaires ask questions, and both expect answers which are true or are in accord with the facts, there is enough difference in their function to permit a differentiation in terminology. In taking a test one *tries* to give the correct answer to the questions. In taking a test, one is aware that he is being tested and bends his energies accordingly; on the other hand he does not *take* a questionnaire—he *answers* it. In answering a questionnaire the issue is not whether a person *can* answer the questions; but whether he *will* answer the questions truthfully. In a test we look to the *difficulty* of the questions and are interested in the speed with which they are answered. In a properly constructed test a person's score is limited by his ability, and he is unable to exceed certain scores no matter how hard he tries. In a questionnaire we eliminate difficulty by making the questions as easy and simple as possible and give the person as much time as he wishes to answer. We place the emphasis on truthfulness of response. In answering a questionnaire one may alter his answers at will, allowing them to portray one or another picture of the situation* to suit a particular purpose. Tests, in short, are designed to find out what a person *can* do, while questionnaires are designed to find out what a person *has done* or *will do*, or how he *thinks* or *feels* or *believes*.

Questionnaires are instruments of research of great value in the

*This latter statement is not unreservedly true. As a matter of fact most persons probably would not be able to go against habit sufficiently to make very extensive alterations to their answers to a questionnaire. The truthful person finds it very difficult to be untruthful.

investigation of conduct. Rugg (32) has classified them into three groups: (a) those asking for facts which the reporter has observed, (b) those asking for facts to be found in records, and (c) those asking for reactions of the individual, such as beliefs, preferences, likes and dislikes, wishes, judgments, and choices. Examples of the first type of questions are "What is your name?" "What is your sex?" "What is your age?" "Where were you born?" "What is your occupation?" "What is your salary?" In this first type all the questions relate to the personal history of the person reporting on the facts, and he draws the answers directly from experience. In the second type, the questions cannot be answered from memory or experience, and usually one must go to records to find the answers. In this group Rugg places questions regarding the age-grade distribution of pupils in a school; distribution of teachers' time among various subjects; statements from payrolls or class enrolment records, etc. In the third type, questions are asked concerning a person's likes, dislikes, beliefs, wishes, tastes, interests, and preferences. The first two types are objective in the sense that the questions pertain to facts which may be verified by others. Witnesses may be called in to identify a man's name, or the records may be resorted to for verification as to age. The last group are subjective in the sense that the answers refer to inner states of the individual which only he is able to examine and observe, and which he alone is in a position to divulge. The accuracy of answers to questions in the first two types of questionnaire is generally taken for granted even though the answers may not in every case be exactly true. However, the psychology of testimony has revealed to us the fact that a man is often unable to observe correctly and often is impelled by motives which cause him to answer incorrectly, and much of the discussion in what follows endeavors to give directions for the construction and use of the questionnaire so as to minimize the factors which lead to untruthful answers.

The third type has been shunned for many years because it was believed that the answers to the questions were so subjective as to be of little value. The renaissance of this type of questionnaire proceeds on an entirely new set of assumptions regarding its use. It is recognized that the answers are subjective and may possess little basis in fact. But the high reliability which these questionnaires possess show that they do measure something quite

consistently. Consequently, while answers to individual questions are recognized as having little significance, the answers to groups of questions may show important trends. Correlations are used to determine what these trends are. In using this type of questionnaire, though one makes no assumptions concerning the truthfulness of the answers, it has been found that certain answers point to conduct trends in the individual. As an example, it has been shown that high school seniors who state that they do not like to play poker are more studious than students who say they do like to play poker. The answers to this question, however, should not be taken as an index of the extent of poker-playing in a particular school even though, as occurred in the case in point, thirty-eight out of 121 answers were in the affirmative.* There may have been a tendency for the other eighty-three boys to attempt to create a favorable impression. In any case, there was a tendency for the thirty-eight who said they liked to play poker to be less studious than the forty-eight who said they did not like to play poker. Whatever the truth of the answers, this tendency toward non-studiousness on the part of those who admit an interest in poker is important.

It is possible for the answer to a question and the interpretation placed upon the answer to be contradictory. For instance in the Character Education Inquiry, a questionnaire was used called *CEI Attitudes S A*.† Examples of some of the items in this questionnaire are (16, p. 98):

5. Do you always preserve order when the teacher is out of the room?

13. Do you usually pick up broken glass in the street?

29. Do you read the Bible every day?

It was found that children who gave *yes* answers to those questions and answered the other thirty-three questions in similar manner tended to cheat *more* in class-room situations, in athletic contests, and with money than children who answered them in the reverse manner. Consequently these tests are considered good tests of lying. The correlations, not the apparent value of the questions, tell the story.

* Symonds, P. M., "A Studiousness Questionnaire," *Journal of Educational Psychology*, 19:152-167 (1928).

† Hartshorne, H., and May, M. A., *Studies in Deceit* (The Macmillan Company, 1928), pp. 98 ff. By permission of the Macmillan Company, publishers.

One should always know whether a questionnaire is being used to obtain facts or to study individuals. Questionnaires of the first two types are used to obtain facts, and the answers are to be taken at their face value. In questionnaires of the third type the purpose is not so much to obtain facts as to study individuals. There may be the wish to investigate the study situation of a school by means of a questionnaire. To find the facts regarding study, one may ask such questions as "How long do you study outside of school each day?" "Do you take notes during your reading?" "Do you review the lesson of the previous day?" To find out how studious an individual is, the writer found the following questions most significant. "Do you like to day-dream?" "Would you like to own a revolver?" "Would you care to become an aviator?" The answers to these need not be taken at their face value as facts, but taken in combination they do show important trends in an individual.

Questionnaires Designed to Elicit Facts

Experience in the use of questionnaires as instruments of investigation has resulted in certain rules and certain warnings which by helping to avoid pitfalls result in more satisfactory returns.

In the first place the questionnaire should be used only in the pursuit of new or original inquiries. Answering a questionnaire is a demand upon a person's time and attention. One must be careful not to misuse questionnaires with any group of people so that resentment or hostility to the method is aroused. There are two exceptions to the general rule that questionnaires be used only in a new or original investigation. One is when there is suspicion that the results of a previous similar investigation are unreliable, the other when after a lapse of time conditions have so changed that a repetition of the inquiry could shed valuable light on movements or trends.

A second requirement is that the questionnaire should be used only in an attempt to elicit vital information. Too often this valuable instrument has been used in picayunish or trivial inquiries. Here one must rely on judgment as to the value of the information sought. One criterion of this value is the sincerity with which it is sought. In recent years the questionnaire has been resorted to by students in pursuit of graduate degrees when the only excuse

for the inquiry was the hope that it would contribute to the dissertation. A strict censorship over questionnaires used for such extraneous purposes would be welcomed by their victims.

In general an inquiry should be pursued only when the results are of value not only to the persons to whom the report is addressed but also to those who help furnish the information. The returns on a questionnaire are always more complete if those to whom it is addressed are made to feel its value. They should be told where the final report will appear; and provided the investigator is in a position to carry out his pledge, those who answer should be promised a personal report on the outcome of the investigation. Since no inquiry should contribute merely to the selfish personal interests of the investigator, it goes without saying that professional advice should never be sought in a questionnaire.

Before embarking on the active prosecution of an inquiry, the investigator should familiarize himself thoroughly with all previous work that has been done on the problem. A systematic search should be made for previous studies that have been made in the same field, and these studies should be carefully read and the findings organized. There is a technique of bibliographical research (which cannot be described here) which should be followed to make sure that previous work on the problem has been utilized. An investigator should be familiar with the known and the unknown, and also with the pitfalls and difficulties waiting for him in his chosen field. The *known* should be separately analyzed and organized and should find no part in the questionnaire unless for purposes of verification.

In this connection it cannot be too strongly emphasized that information asked for on a questionnaire should not be available elsewhere. Only too often the questionnaire is a lazy man's method of obtaining facts or answers which might be looked up in already existent printed reports. Rugg (32) gives as an example the following question included in a questionnaire addressed to teachers: "State the population of the village, town, city, or district in which you teach." Besides the objection that answers given to such a question are unreliable, the investigator himself had access to this information in the census reports.

Furthermore, questions should not be asked which require extensive inquiry on the part of those to whom they are addressed. For instance, it is hardly fair to address a questionnaire to a

school principal for information concerning his teachers—their training, experience, conditions of work, etc.—which would oblige him to canvass his teachers for these facts, when the teachers might be addressed directly by the inquirer. Likewise an investigator has no right to ask for an extensive compilation of records in an administrative office. If he needs these compilations, he should provide for the necessary labor. In certain cases, however, questionnaires may justifiably be sent out which do ask for compilation on the part of those answering. The United States Office of Education, for instance, has asked for compilations in its work of gathering school statistics. Frequently large coöperative studies canvass the members of an association who are severally willing to assemble such data as constitute their share in a larger program in which all are interested.

One should always consider whether the person addressed has the information sought for. It is a common error to address questionnaires to persons who are no more competent to answer them than is the investigator. Questions put to children should be especially scrutinized in this respect. For instance, in some of the questionnaires designed to give a measure of socio-economic level, questions are asked which no child should be expected to answer. "How many books in your home library?" is certain to lead to wild estimates on the part of most children. Even if a little help is given by telling how many books usually go in a foot of space on the library shelves, most children will have as much difficulty estimating the number of feet of library shelves as the number of the books themselves.

Defining the problem. A survey of the literature dealing with previous work on a problem is usually most revealing. What originally appeared to be a single and modest problem will take on many aspects. Questions crop up on all sides. At the earliest opportunity the problem at hand must be delimited, issues must be sharply drawn, and those that are pertinent to the main problem must be separated from those that are to be considered collateral. There is every reason for laying careful plans for every step of the work at the very outset. Just as the architect plans every detail of the construction of a house on paper before a sod is turned, so the investigator ought to foresee all issues that are likely to appear in his work and plan deliberately either to neglect them or to meet them adequately.

128 Diagnosing Personality and Conduct

A novice is apt to underestimate the labor of drawing up, sending out, and tabulating the returns on a questionnaire. Rugg states that the scope of the inquiry must be planned from the beginning. In certain inquiries a complete canvass is desired, in which case the decisions at the outset of the inquiry particularly concern the number and character of questions to be asked. However, most investigations need not aim at a complete census or inventory. If a picture or summary of conditions is wanted, the same results can be obtained by sampling, provided care is taken to ensure representativeness by taking into account all factors which might influence the results. Do you wish to survey practices in the teaching of English? Then be careful to recognize such factors as size of school, section of the country, and types of courses, and plan the investigation to cover all the variations which may occur because of such conditioning factors. "Furthermore, the delimitations of the extent of the study call for careful weighing of the relative value of having a small number of questions and a large number of replies, or of having a large number of questions with a small number of replies." (32) All of these matters must be decided before active work on the investigation is started.

Administration of the questionnaire. It is decidedly a good plan in preparing a questionnaire study to try it out in a preliminary form. The scrutiny of prior studies shows, perhaps, the questions which one wants to ask, but there are many problems connected with formulating the questions and estimating the kinds of answers which will be received which only a preliminary try-out can decide. If it is not possible to try out a preliminary draft on members of the same group to whom the perfected schedule will be sent, try it out on students or colleagues, or even on friends or members of the family. You will find that some of the questions are ambiguous or are easily misinterpreted, and that others are not answered as you intended. In the light of these deficiencies and the criticisms of your friends, revise your blank for its final form.

When possible, mail all schedules in duplicate. If the inquiry is of value to those to whom it is addressed, they will appreciate a duplicate set of their answers, or at any rate a set of the questions, for filing away.

A space should be left at the end of the questionnaire for the

signature of the respondent. But because he may carelessly fail to sign his name, the sender should type in on the question blank the name and address of the one to whom it is sent, thus guaranteeing that these essential data will be on the blank when it is returned.

A *self-addressed*, stamped envelope should be included with the questionnaire as a courtesy.

Buckingham urges that returns on a questionnaire be acknowledged, an additional courtesy well worth extending and one which perhaps can be satisfactorily accomplished by means of a postal card.

It must be recognized that 100 per cent returns from a questionnaire are never obtained, even in the case of questionnaires sent out by government bureaus which have a maximum of prestige and authority. Where conditions are not so favorable, the returns may be no higher than 30, 40, or 50 per cent. Factors which tend to raise the percentage of returns are the authority or prominence of the sender, the attractiveness of the question blank, the pertinency of the topic to the interests of those who are to answer it, and the various psychological appeals adopted to stimulate an answer. Unanswered questions blanks are not always thrown in the waste-basket since oftentimes they are put aside till a more convenient time for answering and thereupon become lost in papers on the desk. Carelessness rather than lack of interest may account for many failures to make returns. It should be remembered that in most inquiries the choice of those to whom schedules are sent is a delicate issue. When a considerable number fail to be returned, the representativeness of the sampling may be seriously deranged. The investigator must always ask himself if there is a probability that those who failed to answer would have given the same answers as those who did. If those who do not answer fail to do so because of lack of interest or because of ignorance, there is strong probability that their returns, if they could be obtained, would alter the results. Usually, when returns are received from only a small percentage of those to whom question blanks are sent, some estimate must be made which allows for this factor of selection.

Toops (36) suggests the use of follow-up letters to obtain a higher percentage of returns. Lindsay (22) found that a follow-up card and letter were of greater value than the original letter and

questionnaire. The usual experience is to find that there are diminishing returns from successive follow-up letters. Toops found that it took six follow-up letters, each using a different type of appeal, to get 100 per cent replies in a certain inquiry, where the returns to the original questionnaire were 52.7 per cent. He found that those best known to the sender were the first to reply, indicating that acquaintance is one factor in producing returns. It is wise in outlining an investigation to plan to include two or three follow-up cards or letters, since these have a demonstrated value.

Preparation of the question blank. Every questionnaire should have accompanying it a letter stating the purposes of the inquiry and enlisting coöperation. Although in some cases a statement performing the same function may introduce the questions in the blank, the use of the letter is probably preferable. The whole success of the investigation largely stands or falls on the strength of this original appeal. Some letters are sentimental, some banter, some cajole, some threaten, some are enthusiastic, some are matter-of-fact. The present writer prefers one that is straightforward and *truthful*. Let it show enthusiasm by all means; but, more important, it should plainly and succinctly state the problem and its importance, both in general and to the person answering. Poffenberger (29) suggests that the preliminary letter be made to play upon deep-rooted motives and desires. One such device which may seem coarse, but which is undeniably effective, is to appeal to the person who is being questioned as one of a select group whose judgment is especially respected. If a questionnaire is addressed to "a leading teacher of experience" or to "one of the more thoughtful parents" or to one whose "experience and judgment is especially valued," this is playing on a motive that few are unresponsive to. Because the appeal of authority is also powerful (the census, backed by the authority of the law, has power to require an answer), questionnaires sent out by institutions or associations are almost certain to command greater respect than those sent out by individuals. A student will always find that his questionnaire receives a larger percentage of answers if his letter includes a statement that his inquiry is sponsored by his instructor, especially if the instructor has a wide reputation. We all feel more inclined to answer questions asked by those whom we know.

The design of the question blank is most important. Its size is

usually determined by exigencies of printing or mimeographing, and of mailing. But convenience in answering and in tabulating and perhaps in filing are also factors. A single sheet of questions is much to be preferred to two or more pages if the questions can be crowded into that compass without making the blank difficult to use. In the matter of arrangement, probably the most serious error is that not enough room is left for the answer. By all means do not cramp the spacing of the questions where an opinion or evaluation is asked which may take several lines. Plan to have answers come at the right-hand side of the page and if possible in a column. The plan of blocking off questions

1	2	3
4	5	
etc.		

irregularly over the page as shown in the illustration is not to be recommended because questions are often overlooked and the tabulator finds that his difficulties are increased.

Some persons may prefer to answer a questionnaire on the typewriter, and the space for answers should be so planned as to make this possible without cramping the room for written answers.

Chaddock (9) warns that paper which will take ink should be used for the schedule. Another good suggestion made by Chaddock is to number or letter each item so that the tabulator may readily identify it.

Rugg (32) advises that the complete tabulation forms be prepared before drawing up the questionnaire. If this is done, the pertinency of each question becomes apparent, and useless questions may be avoided. This preliminary step will also ensure that the investigator words his questions so as to anticipate the form of the answer that can be used best. Often it happens that one wishes too late he had worded his question so that the answer could be tabulated in numerical terms.

He who plans to use the Hollerith tabulating machine or other mechanical methods in tabulating the returns to the questionnaire should confer with the operator of the machine before draw-

ing up his question schedule. From a carefully planned questionnaire, with questions specially printed and numbered in a certain way, an operator can transfer the data to the Hollerith cards much more economically than from the usual form of questionnaire blank.

Finally, the investigator must get his question blank printed or typed. Since the attention which a questionnaire receives is directly affected by the attractiveness of the form in which it is put out, he should spare neither care nor expense in preparing the blank. Whenever possible have it printed, as a printed sheet is in the cleanest and most attractive form. Multigraphing is better than mimeographing. If the mimeographing process is employed, use a good quality of paper, and keep the copy clean by cutting fresh stencils when necessary.

Preparation of the questions. The following rules have been found helpful in preparing questions for the schedule of an investigation.

1. *The number of questions should be small.* It is a temptation to ask many questions. As one plans an inquiry, its scope tends to enlarge, and there is a tendency to allow the original problem to spread into by-paths. One argues that since he is canvassing this particular group of people, he may as well ask this or that question, interesting in itself, though not strictly pertinent to the inquiry. The mass of questions rolls up like a snowball. This tendency toward expansion must be combated by rigorous censorship. No fixed standard can be set for determining what is the smallest possible number, because each inquiry has its own conditions, but a maximum of ten questions is probably safer than one of forty.

2. *Questions should be brief.* Questions should be worded in the fewest possible words. Probably the questions stated most briefly also excels in clarity and directness. Poffenberger (29) asked seventy-five persons to state a question to find out how much people usually pay for their socks (or stockings). Seventy-five different questions were framed—no two were alike. They ranged all the way from "What do your socks cost you per pair?" to "We are trying to get information concerning the average price paid for hosiery. Will you please tell us what you pay for your hosiery, specifying the kind of hose you wear, socks or stockings, and the make?" One has only to imagine the form in which the answers to

the longer question will come to see how difficult they will be to use.

3. *Questions should cover information desired.* Frequently a question is so worded that it does not ask what is intended. In a questionnaire to determine health practices in a school, such questions as "How much milk should you drink a day?" or "Do you usually sleep over eight hours a day?" are quite inadequate to get the information desired. Instead of asking a person what he does, the first question asks for a judgment as to what should be done. In the other case the question will merely call forth *yes* and *no* answers without making it possible to determine the average number of hours or the distribution of hours spent in sleep.

4. *Questions should be simple enough to be understood.* A questionnaire is not an intelligence test. In fact, it defeats its own purpose if it is not so understandable that the most unintelligent person to whom it is addressed can answer the questions. One way in which this point can be tested is to make sure that the vocabulary of the questions is not difficult, perhaps by checking the words used to see that they come well down on the Thorndike Word List. Substitute a simple word for a more complex word where possible. Avoid technical words where a common word will do as well. If technical words must be used, which might be unfamiliar to any member of the group addressed, they should be defined.

5. *Questions should be unambiguous.* It is startling to discover how difficult it is to make a question which will not be misinterpreted, or better *cannot* be misinterpreted. In a questionnaire on study it was asked "What kinds of notes do you take?" Such a question may have the greatest variety of answers, for there are so many different categories of notes—long or short, running or outlined, etc.—that the replies would be difficult to summarize.

6. *Questions should be specific, not general.* The general question is the greatest offender in questionnaires. One finds at every turn questions which ask for opinions, but which, failing to define the form the answer should take, yield useless answers. Rugg (32) gives examples of such questions:

Do you have difficulty in obtaining clerical help?

What in general is the attitude of the parents toward "home work" in school studies?

134 Diagnosing Personality and Conduct

What differences in training do you notice between public high school commercial graduates and graduates of the common business college?

By means of questions of this type one can justify the place of almost any subject in the school curriculum. Recently the Classical Investigation canvassed Latin teachers as to the validity of certain objectives for the Latin course in high school. Of the Latin teachers canvassed, 93 per cent stated that "the development of certain desirable habits and ideals which are subject to spread, such as habits of sustained attention, orderly procedure, overcoming obstacles, perseverance; ideals of achievement, accuracy, and thoroughness; and the cultivation of certain general attitudes such as dissatisfaction with failure or with partial success," was a valid objective for the high school course. This answer suggests the testimonials that are often given to discharged employees. While the good faith of such an answer need not be suspected, we are left wondering to what degree teachers believe this ideal is accomplished.

Questions like, "Is the candidate well educated?" or "Do you consider Mr. Blank an experienced teacher?" are practically worthless.

7. *Questions should be stated in acceptable language.* Double negatives or alternatives in the question should be avoided. The following question, cited by Poffenberger, is an example of a badly worded question: "When buying face powder, do you or do you not usually consider whether or not it is of a very fine texture?"

8. *Questions should be so arranged that the answers can be made by checking.* Before inserting a question in the blank, one should anticipate the possible answers. If the answer is to be *yes* or *no*, it is well to put in parentheses after the question, or in small print beneath the space for the answer ("write yes or no"). In some cases "yes no" is printed in the blank so that the answer chosen may be indicated by underlining. This is apt to lead to confusion as some people will cross out the answer they do not consider correct, making it difficult to tell whether underlining or crossing out was intended.

In case the answer is not a mere alternative such as *yes* or *no*, all possible answers should be anticipated and listed in the questionnaire, the answer to be indicated by checking the alternative selected.

Example

Why did he leave school?

Moved away.

Poor quality of school work.

Poor health.

Must work to help support family.

Lack of interest.

Parents did not see value in education.

Low ability.

In this example the correct alternatives should be checked. Often the investigator will not be able to think of all possible alternatives. His preliminary use of the questionnaire on a small group can then be employed in getting typical responses which may be inserted for checking in the main inquiry. Even with this method, not all possible answers will be listed, and blank spaces should be left for additional answers. It is a fact, however, that such additional blanks are seldom utilized. It would seem as though, for most people, the suggestions made by the possible answers listed in the questionnaire block the recall of other answers.

In case the answer to a question is a series of numbers, the question blank should provide a tabulation blank where the answers may be conveniently listed. Rugg (32, pp. 54, 55) gives the illustration shown on page 136.

This general plan of anticipating possible answers and listing them for checking, or of providing forms in which answers may be conveniently written, helps both to make the questions easier to answer and to make the answers easier to tabulate. Occasionally one deliberately asks a question when freedom in answering should be allowed. There are times when the investigator wants a free expression of opinion, such as a statement as to the success or value of an undertaking, or when he is looking for a variety of suggestions, perhaps for future work. For answers to such questions, one usually needs to allow a half or even a full page for the answer.

9. *Avoid leading questions.* Questions should be so worded that they do not suggest an answer, or suggest one answer rather than

136 Diagnosing Personality and Conduct

Omitting all names, will you give the individual yearly salaries that were paid junior high school principals, teachers, supervisors of special subjects, and principal's clerk during the school year 1914-1915?

To make it easier for you, the salaries are arranged in groups in one column. In the opposite column (marked "number receiving"), will you place the number who received the salary stated?

Example—If four women teachers and one man teacher receive *annual* salaries between \$800 and \$825 respectively, enter them thus:

<i>Annual salaries</i>	<i>Men</i>	<i>Women</i>
\$800-825	1	4

<i>Principals' salaries</i>	NUMBER RECEIVING EACH SALARY GIVEN		<i>Teachers' salaries</i>	NUMBER RECEIVING EACH SALARY GIVEN	
	<i>Men</i>	<i>Women</i>		<i>Men</i>	<i>Women</i>
\$1000-1099			\$ 500- 549		
1100-1199			550- 599		
1200-1299			600- 649		
1300-1399			650- 699		
1400-1499			700- 749		
1500-1599			750- 799		
1600-1699			800- 849		
1700-1799			850- 899		
1800-1899			900- 949		
1900-1999			950- 999		
2000-2099			1000-1049		
2100-2199			1050-1099		
2200-2299			1100-1149		
2300-2399			1150-1199		
2400-2499			1200-1249		
2500-2599			1250-1299		
2600-2699			1300-1349		
2700-2799			1350-1399		
2800-2899			1400-1449		
2900-2999			1450-1499		
3000-3099					

another. Münsterberg long ago gave an apt illustration (cited by Poffenberger, 29, p. 139) of the leading question:

"When a clerk in a store says to a customer: 'Will you take the package with you?' an affirmative reply is anticipated and frequently received. To the question 'Will you have the package sent?' an affirmative reply is again anticipated and frequently received." *

* Poffenberger, A. T., *Psychology of Advertising* (A. W. Shaw Company, 1925). By permission of the present publishers, McGraw-Hill Book Company.

So we see that such a simple thing as the form of questions asked by clerks may mean dollars and cents to a firm.

Muscio* studied the influence of the form of a question on suggestibility. He presented moving pictures as material for observation and asked a number of questions stated in eight different ways. The suggestiveness of a question was measured by determining the percentage of times a subject followed the lead of the question. The questions in order of suggestiveness are:

TABLE 20
PERCENTAGE OF TIMES A SUBJECT FOLLOWED THE LEAD OF A QUESTION
(from Muscio)

	% suggestiveness
1. Didn't you see a?	91.7
2. Did you see a?	89.2
3. Didn't you see the?	84.0
4. Was the (k) <i>m</i> or <i>n</i>?	77.5
5. Did you see the?	62.6
6. Wasn't there a?	51.8
7. Was there a?	43.6
8. Was the (k) <i>m</i>?	39.7

The negative form of question is always more productive of suggestibility than the positive, and a question which implies a personal relation of the observer in seeing is more productive of suggestion than a mere question of fact. The definite article, *the*, aroused suggestibility less than the indefinite article *a*.

10. *Ask questions that will be answered.* It is common sense to assume that incriminating questions will not be answered, or that if they are answered, there is no assurance of the truthfulness of the answers. Incriminating questions regarding health, religion, economic status, and habits are frequently asked in such blanks as the application for a marriage license, driver's (automobile) license, teaching position, etc. Questions concerning habits in respect to use of intoxicating liquors were evaded even before the prohibition amendment. College students hesitate to incriminate themselves by stating the amount of time they spend in study or the amount of help they receive from others. Questions concerning all matters considered private, such as one's income, or matters which are the subject of moral scrutiny, and social taboo questions

* Muscio, B., "The Influence of the Form of a Question," *British Journal of Psychology*, 8:351-389 (Sept., 1916),

are likely to go unanswered. If they are answered, the answers cannot be trusted.

A different type of question that will not be answered or should not be, is one concerning which the person questioned has no information. Questions that relate to incidents or details wholly outside one's range of interest and attention will not receive answers. Try yourself to see if you can recall the number of pickets in a certain fence, or the number of steps leading up to a familiar building, or the number of pillars in the portal of some church. Although you have passed these buildings daily, you may never have chanced to make these particular observations. Yet many of the questionnaire items require precisely similar detailed observation.

Another type of question which will not in the nature of things receive a correct answer, and perhaps no answer at all, is one which delves back far into the subject's past experience. This particular type of question is now receiving much attention because of the claims of the psychoanalysts that they have thereby discovered in such questions the clues to much pathological conduct. But our knowledge of the learning process makes it clear that most of our established ways of acting are not the product of a single incident but rather of a long series of similar incidents. Learning is not cataclysmic; it requires many repetitions of the same act, and hence an inquiry into the past to determine when one first did so-and-so is almost certain to result in inaccurate answers.

Better than to ask for percentages in a question is to get the data from which the percentages may be computed. Every one cannot be trusted for accuracy in computation, and there is less liability of error in interpretation if one asks for the raw data. Rugg warns us not to forget to ask for the total number of items in case a percentage is desired. For example, an investigator who wished to know the *proportion* of brothers and sisters who lived to adulthood; asked: "How many brothers and sisters lived to early adulthood or longer?" and omitted to ask for the *total number* of brothers and sisters. This particular question is also faulty in that adulthood is not defined and hence will not be interpreted in the same way by every one.

The Psychological Questionnaire

The third type of questionnaire which asks for reactions of the individual we have termed the *psychological questionnaire*. Whereas the first two types ask for questions of fact, the purpose of the psychological questionnaire is to study the individual by asking questions relative to his beliefs, wishes, likes, interests, tastes, preferences, choices, feelings, fears, and worries.

These questions may be divided into two groups according to the way in which it is intended that they be answered. On the one hand instructions may be given that the issue is to be carefully and deliberately weighed and the answer given only after reaching a thoughtful decision. On the other hand one may instruct a person to give his immediate impression, or choice, or leaning, thus aiming to make the answers less a result of careful ratiocination and more of those sets or readinesses that determine our immediate responses to various stimuli. This latter is the type of response that usually is desired in psychological questionnaires. Many questions are asked, and they are to be answered in a short time. Though some have spoken of this as being an emotional response, perhaps in contrast with a thoughtful response, it seems doubtful whether the emotions play more than a minor rôle, if indeed they play any, in determining the answer. The emotions, taken strictly by themselves, are not directive. The response which is wanted in these questionnaires is the resultant of accumulated past experiences, rather than that of immediate ratiocination.

Persons interested in the measurement of conduct have not been slow to adopt the testing devices previously found useful for testing judgment in the school subjects. Goodwin Watson* describes various adaptations of these tests for measuring attitude. These new-type questions aim at objectivity by suggesting possible answers from among which the pupil is to select by checking the answer which he thinks is most satisfactory. Various classifications of these questions may be made. Here they will be merely listed in order and the reader, by noting likenesses and differences, may build up his own classification.

* Watson, G. B., *Experimentation and Measurement in Religious Education* (The Association Press, 1927).

1. *True-false.* In this type a statement is given which may be either true or false, and the answer is indicated by writing T to indicate "true" or F to indicate "false," (or + or -) opposite the statement.*

Example.

I brush my teeth daily

True False

I bathe daily

True False

2. *Yes-no test.* This test, composed of questions to be answered by either *yes* or *no*, is already in frequent use in psychological questionnaires. It is similar to the true-false test, with the statement in this form turned into a question.

3. Goodwin Watson** makes considerable use of true-false or yes-no test items which permit a third alternative answer indicating that the statement is neither true nor false or is to be answered by neither a categorical *yes* nor a *no*. This considerably broadens the scope of statements or questions which may be used in the questionnaire, and more discrimination is needed to decide between three alternatives than between two.

4. The *multiple-choice item* has had extensive use in achievement testing, but less is known concerning its value for use in the questionnaire. The multiple-choice item is a statement which ends with several alternatives.

Examples.

I bathe daily, several times a week, every week, less frequently than once a week.

I brush my teeth after every meal, after breakfast and before going to bed, once a day before going to bed, infrequently.

* Manuals in which detailed directions for constructing true-false and other types of objective tests are given have been prepared by the following:

Hopkins, L. T., *The Construction and Use of Objective Examinations*, (1926), University of Colorado Bulletin, 119 pp.

Odell, C. W., *Traditional Examinations and New-Type Tests* (The Century Co., 1928).

Paterson, D. G., *Preparation and Use of New-Type Examinations* (World Book Company, 1925).

Ruch, G. M., *The Improvement of the Written Examination* (Scott, Foresman and Company, 1924).

Ruch, G. M., *The Objective or New-Type Examination* (Scott, Foresman and Company, 1929).

Russell, C., *Classroom Tests* (Ginn and Company, 1926).

Symonds, P. M., *Measurement in Secondary Education* (The Macmillan Company, 1927).

** Op. cit.

Taken as matters of judgment, the distinction between the questionnaire and the rating scale disappears. The answer to this kind of a questionnaire is a form of rating, and if one answers the questions concerning himself, it becomes self-rating. The question frequently arises as to who should be the person to answer a questionnaire on personal matters—the person himself or one of his friends. The factors which help decide this issue are discussed under self-rating in Chapter III. The questionnaire may even take the form of a rating scale. Instead of having the questions answered *yes* or *no*, relative values may be indicated as in rating. The graphic rating scale makes a very convenient form for the questionnaire when a graded answer is desired.

Looked at from this angle, any question may be answered by degrees on a scale. For instance, in an interest test the following scale was prepared for use in checking the answers to each question: L! L? D D!, L stands for like and D for dislike. The exclamation point was used to denote warmth or certainty of feeling, while the interrogation point denoted a neutral position. Hart* in his Test of Social Attitudes employed three symbols, +, —, to accompany each item. The + was to be encircled to indicate liking, the — to indicate dislike, and the period to represent neutrality. Further, the items were printed in groups of fifteen to seventeen, and in any group five items were to be underlined to denote those items about which the person answering felt most strongly. Finally one item was to be chosen from each group about which feeling was strongest of all. The writer believes it preferable, however, to provide a rating scale for each item so that extreme opinions or feelings may be expressed in the case of more than one item. While it is probably true that one item in each group is reacted to with more feeling than any of the others, at the same time it is unwise to define the number which shall receive extreme ratings.

In this connection it should be emphasized that in psychological questionnaires *every item should be answered*, so that all may receive statistical treatment. In case the item is unknown to the respondent, he may choose for his answer the neutral or average position. The interpretation placed on scores of questionnaires of this type depends on the position of the score within a distribu-

*Hart, H., "A Test of Social Attitudes and Interests," *University of Iowa Studies in Child Welfare*, Vol. II, No. 4 (July 1, 1923).

tion determined in some typical group. In order that the comparison may be valid, the same number of items (all the items in the questionnaire) should be answered. If this is not done, some kind of percentage must be used, and the interpretation of a percentage is not certain when the answered items are carefully chosen and hence are probably not of equal value.

The direct vs. the subtle question. Some of the most successful questionnaires designed to measure attitude have used a disguised form of question. Children in answering questions seemingly intended to elicit information concerning their beliefs or preferences are thereby made to show you important behavior trends in other directions. The Watson "Survey of Public Opinion on Some Religious and Economic Issues" contains good examples of this.* In taking these tests the subject makes certain choices or judgments, thereby disclosing his tendency to radicalism or liberalism in economic or religious fields. One of the tests in the Watson battery gets at these attitudes by determining the tendency toward rationalization. In one test the subject is asked to judge the strength of arguments, thereby disclosing his tendency to disregard the logic of the situation in favor of a prejudice in one or another direction. In another test a situation is briefly described and certain inferences are drawn. The subject is required to judge whether the conclusions fairly follow or not, thereby giving himself an opportunity to show prejudice in favor of one or another side illustrated by the inferences. In still another test the subject is required to decide between the same issue when it is immediate and personal, and when distant and impersonal. The degree to which the subject is inconsistent becomes a measure of the direction and degree of prejudice.

The present writer in his "Studiousness Questionnaire" ** tells classes to whom the test is given that the purpose of the questionnaire is to discover children's interests. Children never and teachers seldom surmise that the purpose of the questionnaire is to measure the tendency toward studiousness.

It is probable that disguised questionnaires are more valid than those which are straightforward in their approach. The straightforward attack partakes too much of the nature of a test and per-

* Watson, G. B., *The Measurement of Fairmindedness*, Teachers College Contributions to Education, No. 176 (1925).

** *Journal of Educational Psychology*, 19: 152-167 (1928).

mits the pupil to control his responses to fit his purposes. The disguised questionnaire, in which the pupil is told he is doing one thing, but in which the items are so selected that the result yields a measure of something else, is the ideal situation for measuring conduct. Since a measure of conduct should be a record of responses in a prescribed situation, the test or questionnaire itself often becomes a factor sufficiently extraneous to the situation to spoil the results for generalizing purposes.

Can children introspect? When conduct questionnaires are given to children, the investigator should ask himself the question, "Can children introspect well enough to answer these questions?" Introspection is a particularly difficult type of observation and requires considerable practice. Children do not naturally observe their own habits or methods of work and play. Yet we assume in asking our questions that they have already made observations of their conduct and are quite ready to answer our questions.

This form of retrospection, of looking back over one's behavior processes and recounting them, is a feat that requires special training. To make this more evident let the reader try to answer the following questions concerning golf:

Do you usually take one, two, three, or more preparatory swings before driving?

Does it bother you to have some one watch you drive?

Do you usually drive harder on the first or last hole?

How much time (percentage of total) do you spend looking for lost balls?

How long as a rule do you spend in building a tee?

Do you look at your hands, the ball, the club, or the hole in putting?

How do you prepare for a match?

And yet these questions do not differ in character from the following which were gleaned from one of the questionnaires used to obtain data from high school students as to their methods of study.

Do you usually read the assignment once, twice or three times?

Does it bother you to study in a room where other people are talking?

Are those lessons studied at home your hardest or easiest?

How long, as a rule, do you spend on each lesson in the subjects you are carrying?

How long do you study at a time?

Do you read your lessons aloud? Do you close your book and repeat your lesson to yourself?

Do you try to think out your lesson mentally just before going to class?

How do you prepare for an examination?

These golf questions seem to a golfer ridiculous. Do the study questions seem ridiculous to a student? The only golfer who stops to analyze his movements is the golf instructor or the professional who wants to write a book.*

Probably the answers to questionnaires which require introspection by young children are largely chance answers. To be on the safe side, one should either require that children observe their own conduct or habits of work for a week or so before the questionnaire is used, or the questions should cover familiar facts and everyday experiences which are likely to be observed.

Can adults retrospect to their childhood? Another pressing problem with regard to the use of the questionnaire method for the collection of facts of conduct is whether adults can accurately recall facts and experiences of their childhood and adolescence. Years ago Thorndike ** answered this by saying "Adults even so well trained as college seniors and even in the simplest matters of present objective fact such as are involved in the questions, 'How tall are you?' and 'What is the circumference of your sister's head?' make gross errors. The errors increase in number and amount when the report requires memory; increase further when the fact is a report of subjective condition; and multiply like bacilli when it involves the consideration of the general drift of a series of experiences."

And yet the whole theory of psychoanalysis is based on belief in the integrity of recalled personal experiences. Breuer's "Cathartic Method," which is the basis of Freud's later method, is described by Wohlgemuth as follows: "A highly unpleasant experience occasions a psychical shock, or wound, or trauma, which gives rise to certain reactions. The occurrence itself is forgotten and as far as consciousness is concerned, is apparently obliterated, for it cannot be recalled or revived by ordinary means. It, the

* Symonds, P. M., "Methods of Investigation of Study Habits," *School and Society*, 24:146-147 (July 31, 1926).

** Thorndike, E. L., *The Original Nature of Man* (Bureau of Publications, Teachers College, 1913), p. 32.

memory of the occurrence, continues, however, to exist unconsciously, it is jammed in, and acts like a foreign body in a wound. It manifests itself by the continuance of the original reactions, which thus become hysterical symptoms. Bringing the memory of the occurrence back into consciousness is like extracting the foreign body from the wound. The emotion, which at the occurrence of the shock had had no opportunity to 'work itself off,' and so was jammed in, was strangled, can do so now, and the hysteric symptoms cease to recur." *

Freud's theory of the importance of infantile reactions determining subsequent behavior has had phenomenally wide notice, and recent attempts to demonstrate the importance of early childhood experiences in causing subsequent maladjusted behavior have taken the report of childhood experiences at face value. Chassell,** for instance, in his "Experience Variables Record," provides sections for noting the occurrence of experiences in *childhood*, *early teens*, and *recent* (times). Chassell recognizes that there is a tendency to remember outstanding events and to neglect the more important habit-forming experiences of everyday life. He says, "An interesting problem remains of how to distinguish between rare events that were, perhaps, strongly influential ('traumata'), and the ever-recurring, but less 'affective' day by day round of events." Later on in giving correlations of reported early experiences with present adjustments, he says, "In considering this evidence, we must not fail to bear in mind that the data are memories and reported attitudes, and that the interrelations are to this extent subjective, rather than being associations between present observed attitudes and proved past events."

But this cautious attitude is by no means universally held. House has to explain the fact that the best adjusted of his college students reported the most maladjustment as children. The obvious explanation that memory of childhood experiences is inaccurate he disclaims: "Though recency may be an important factor in accuracy of recall, frequency and intensity are no less significant for the persistence of memories: as witness the tenacity of

* Wohlgenuth, A., *A Critical Examination of Psycho-Analysis* (Allen and Unwin, 1923; The Macmillan Company, 1924), p. 46. By permission of The Macmillan Company, publishers.

** Chassell, J. O., *Experience Variables: a Study of the Variable Factors in Experience Contributing to the Formation of Personality* (published privately, Rochester, N. Y., 1928).

memories rooting in early childhood and preserved even in senility. We might reasonably argue, then, that being older does not of its own accord militate against the retention and recollection of memories sprung from childhood situations, especially when these latter are universally distributed, and in no sense unique occurrences."*

Our conclusion is that it is exceedingly risky to take memories of childhood experiences at their face value. Even if they were real experiences, their affective accompaniments may have magnified them out of all proportion to their true importance. Probably most of our recalled experiences are recalls, not of the experiences themselves, but of our memories. Even so, as Chassell found, answers to questions as to childhood experiences may have high diagnostic significance. Perhaps this significance is due to the construction which we place on our past experiences rather than on the significance of the experiences themselves.

Questions of reaction vs. questions of environment. Questionnaires designed to measure various types of adjustment may be divided into those which ask questions concerning a person's reactions and those that ask questions concerning a person's environment. The Woodworth "Psycho-neurotic Inventory" illustrates the former. This particular questionnaire asks questions about the subject's worries, fears, physical disturbances, and social adjustments. The Chassell "Experience Variables Record" asks questions of the other type, grouped about such phrases as "Mother's Urgency toward Subject Carrying Out Her Way," "Mother's Severity in Dealing with Subject," "Tendency of Parents to Favor Other Children above Subject," etc. The measure of home environment could also be classed with this group. Accurate comparisons of these two types of questions for measuring adjustment have not been made. Reasoning *a priori*, it would seem as though more reliable results might be obtained by using the questions on the environment, since it seems probable that these matters have been the object of observation more frequently than one's own responses. Individuals differ, however, in their tendency to observe the details of their social surroundings or of their own reactions. Research alone can tell us which is the better type of question to use for this purpose.

* House, S. D., *Mental Hygiene Inventory*, Archives of Psychology, No. 88 (1927), p. 21.

Validity of Questionnaires

One should distinguish carefully between the validity and reliability of psychological questionnaires. Questionnaires, even though somewhat unreliable because of the inaccuracy of the ratings or judgments given as answers to the questions, may nevertheless have high validity. If this is the case, the instrument could be made more reliable by extending it, making it more objective, or standardizing its procedure. Once we have demonstrated that we are on the right track in measuring important attitudes or phases of personality, technical skill can be applied to making the instruments more accurate.

Selection of items. In selecting valid items for a questionnaire there are two methods in vogue. One consists in careful analysis of the field being measured, followed by assemblage of questions on the basis of this analysis. This was the method employed by Heidbreder* and Freyd** in measuring introversion-extroversion. Freyd made a careful analysis of the manifestations of introversion-extroversion, drawing on the writings of Jung and others, and Heidbreder assembled the analysis into a set of questions.

The other method is empirical; the items are selected by trying them out on groups of subjects varying in the quality or characteristic that is to be measured. There are two methods of obtaining this variation in the subjects to be used in validating the test items. One is to have them rated by associates on the quality to be measured. The other is to use groups that have already been selected socially so as to represent extremes of the quality. The latter method, for example, was used by Lentz† in attempting to discover tests of delinquency. He tried out his tests on boys in a probation school as representing the delinquent group and boys in a public school as representing a normal group. Society in this case had already made the selection. Cowdery††

* Heidbreder, E., "Measuring Introversion and Extroversion," *Journal of Abnormal and Social Psychology*, 21: 120-134 (July, September, 1926).

** Freyd, M., "Introverts and Extroverts," *Psychological Review*, 31: 74-87 (1924).

† Lentz, T. F., Jr., *An Experimental Method for the Discovery and Development of Tests of Character*, Teachers College Contributions to Education, No. 180 (1925).

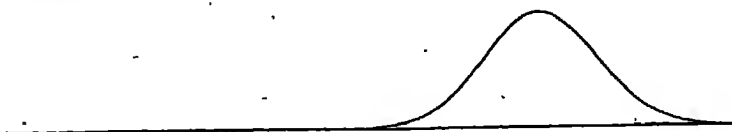
†† Cowdery, K. M., "Measurement of Professional Attitudes," *Journal of Personnel Research*, 5: 131-141 (August, 1926).

148 Diagnosing Personality and Conduct

used the same method in validating his interest questionnaire on doctors, lawyers, and engineers—groups already socially differentiated.

In this second type of item-selection, one makes only a rough analysis of the characteristic being measured, admitting that such an analysis is premature before the characteristic has been studied objectively. To this end, rather a large and miscellaneous collection of items is gathered together which might by chance show some relationship to the quality being measured. These items are then tried out on the subjects representing degrees of the quality in question, and those items are retained which are answered differently by the contrasting groups, or which show a substantial correlation with known degrees of the quality in the criterion group. This was the method employed by Strong in determining the scoring key for his interest questionnaire in differentiating the professions which he studied.

A note of caution must be sounded here. The validity of the questionnaire must be tested out on a *new group*, not on that used in selecting the items. In violation of this rule, remarkably high validity coefficients have been obtained by computing validity on the same data which were used in choosing the items. The original elation over this achievement turned to disappointment when the questionnaire failed utterly to show as favorable results in a new situation. This phenomenon, which we may call "weighting the probable error," is to be explained as follows: The score made by any pupil on a test or questionnaire is the result of



PROBABILITY CURVE

The theoretical probability curve of the distribution of scores made by an individual repeating a test many times. The base line represents the scale of scores of the test.

many factors, some of which are chance factors, so that if the individual were to repeat the test even under similar conditions, his score would not be the same as on the first trial. It is known that if an individual were to repeat a test many times his scores

would form a normal probability curve centering around some value which is his true score.*

Now when items are selected empirically, as desired above, and the resulting test or questionnaire has a reliability greater than zero, if then a score is computed for any individual on the same data, this score will be above or below the true score for that individual. If his score is high, it is too high, and if it is low, it is too low, both in the direction which yields the most advantageous correlation with the criterion. When the test is repeated there is a tendency for these scores which are too high or too low to fall back closer to their true scores, which tends to lower or attenuate the relationship with the criterion. In short, the safest way to avoid this illusory phenomenon is never to compute a validity coefficient on the same group which was used for selecting the items or for determining the weight of the items.

Besides this spuriously high validity which results from trying out tests in the same situation as that in which the items were selected, there is some reason to believe that questionnaire results depend on the special situation in which they were obtained more than do test results. Ability as shown by tests has a remarkable consistency, but the adjustments or interests which are tested by questionnaires tend to be more ephemeral—the ache or pain may have been specially noticeable the last few days; enjoyable activities may be seasonal; one is constantly adding new friends, acquiring new ambitions, or wearing out old fears. We can, then, never hope for the consistency in the case of questionnaires that is found in the best tests. However, this matter may be overmagnified, for on the one hand questionnaires have actually shown a consistency that has surprised even those who have worked with them, and on the other hand these variations need not be deplored, since they probably indicate important and significant changes in the individual.

Veracity. The veracity with which questionnaires are answered is a matter of great importance. There are two cases when questions will not be correctly answered: (a) when the person questioned *does not know* the answer, and (b) when he *does not want to give* the correct answer. Before taking up the rationale of the

*Thorndike, E. L., "The Variability of an Individual in Repetitions of the Same Task," *Journal of Experimental Psychology*, 6:161-167 (1923).

matter of veracity, let us describe the rather meager experimental findings.

Persing (28) made a comparison between what students in a chemistry class say they would do in reporting errors in the grading of examinations returned to them and what they actually do. He systematically made errors in grading examination papers for a given class. Some grades were entered too high, others too low. The students were requested to report to him any errors which they found in the marks on their papers. Grades were reported as too low by 97 per cent, but only 9.5 per cent reported grades as too high. Later in the year Persing asked these same students various questions regarding their attitude toward examinations. The results showed that 97 per cent said they would report low grades if they found them and 80.5 per cent would report grades that were too high. The comparison is between this profession of 80.5 per cent of the students with the actual 9.5 per cent who did so report over-grades. It suggests that statements of attitude are to be taken at face value if they are neutral or in favor of the interests of those concerned, but that statements of attitude tend not to agree with conduct when it is a matter of opposing one's own interests.

Hartshorne and May* during the course of their experiments used a "Pupil Data Sheet" which included, besides a number of questions of a general character, twelve questions relating to cheating on tests in school. Samples of these are (16, p. 127):

33. Did you ever cheat on any sort of test?
34. Have you cheated on such tests more than once?
45. On some of these tests you had a key to correct your paper by. Did you copy any answer from the keys?
57. Have you answered all the questions honestly and truthfully?

These experimenters had previously administered honesty tests to these same school-children to find out to what extent the children would cheat if given opportunity. The answers to these questions were then matched with the record of their actual cheating. A lie index was constructed to measure the extent of their lying, such that if they lied more often than not, the lie index

*Hartshorne, H., and May, M. A., *Studies in Deceit* (The Macmillan Company, 1928), pp. 94 ff. By permission of The Macmillan Company, publishers.

was positive; if they admitted their cheating, the lie index was negative; if the lies and truths were equally divided, the lie index was zero. This lie index was applied only to those children who actually cheated, as "the others had nothing to lie about." In one school it was found that 83.0 per cent lied, 15.5 per cent were truthful and 1.4 per cent showed a balance of truths and lies. In another school 85.1 per cent lied, 10.7 per cent were truthful, and 4.1 per cent showed a balance of truths and lies. The conclusion must be that one does not get truthful answers from children about their conduct when it is against their interest to tell the truth or when they evidently would be criticized for what they said.

In another investigation Wylie (41) studied the validity of answers to a questionnaire given in school to determine the socioeconomic status of the home. This study was a piece of follow-up work growing out of the Newark Survey of the all-year school. Case workers were able to go into the homes of twenty-nine of the children and check up on the answers to their questions as to the home furnishings, occupation of parents, etc.

The percentage of accuracy of some of the items is given in the following table:

TABLE 21
PERCENTAGE OF ACCURACY IN ANSWERING A QUESTIONNAIRE
(from Wylie)

	<i>Percentage of exact agreement</i>
Did mother attend high school?	92.9
Number of books in home library.	24
Telephone in home.	92.9
Family possession of auto.	82.8
Bathtub in home.	82.8
Does mother work?	92.3
Occupation of father.	62.5
Country in which father was born.	100
Country in which mother was born.	88.9
Language mother speaks.	76.9
Numbers of rooms in house.	69.0
Number of persons in family.	58.6
Possession of library card.	24

This follow-up work gives a very good picture of the accuracy with which children answer questions of this type. Matters of fact such as possession of a telephone are answered with a high

degree of accuracy. Questions where some interpretation is necessary or where the item is not immediately observable are answered less accurately. In the case of the father's occupation some children see him leave home every morning but have no idea where he goes or what he does. Others know in a vague way that he works in a factory or office but have little conception of his position or the type of work he does. In other questions it is a matter of definition, which may perhaps account for the errors in giving the number of persons in the family. In still other cases there is a definite lack of observation and we should not wonder at the low percentage of correct replies to the question concerning the number of books in the home. In this case Wylie goes on to show that the percentage of agreement between children and case workers, if a leeway of ten volumes is allowed, is 60 per cent and if a leeway of twenty-five volumes is allowed, 76 per cent.

It has many times been pointed out that one does not get the truth by asking many people who do not know what the truth is. There is no magic in the number of persons who answer a questionnaire. Yet how distressingly often a student, seizing upon a question to which no one knows the answer and which has an important bearing on educational practice, will attempt to gain a solution of it by sending out a questionnaire to be answered by persons who know perhaps even less about the matter than he does. Statistical methods cannot extract the truth by taking the average of prevailing ignorance as to the objectives of different studies, best methods of work, standards of action, and the like.

On the other hand, Wylie's results show that whereas one cannot necessarily take an individual's answers at face value, the total or average of a number of answers may possess a high degree of accuracy. For instance, although only seventeen out of twenty-nine children (58.6 per cent) gave the correct answer as to the number of persons in the family, the total number of persons mentioned as members of one family or another was 202 for the pupils as against 204 for the case workers, representing an accuracy of 99.0 per cent. If errors are compensating, then the total or average may represent a high degree of accuracy. Again, one out of four children answered wrongly as to the possession of a library card, yet the total number of cards reported by the library was 115 as against 131 reported by the pupils, or an agreement of 86.1 per cent. Taken in this sense, the average of a large

number of answers possesses considerable validity, for, statistically speaking, the reliability of the average may be increased by increasing the number of individuals canvassed. Errors tend to compensate when the answer is a pure matter of judgment. Errors tend to be cumulative when there is bias or "interest" in the answers.

It is seldom possible to get at the truth by asking a question to determine desirable norms of conduct when the answers depend on the cumulative experience of those answering it, for habit plays a large part in determining attitudes. Matters of clothing, taste, and style are notoriously conventional, and obviously depend on the conditions of previous experience. One's tastes or preferences should never be confused with the truth of an issue, since often our preferences have little relationship to logic.

Wylie recommends the use of interlocking questions as one device for obtaining greater accuracy from the results of a questionnaire. For example, to ascertain age it is well to ask both for the age and for the date of birth, so that one may be used as a check against the other. In cases where there is a deliberate attempt to deceive or conceal, as must have been the case with some of Hartshorne and May's results, the consistency of the answers would be no check against their truthfulness.

Davis (10), in studying the answers to a questionnaire to children on health habits, deduced the following checks of accuracy:

(a) The answers should check with general knowledge and experience. If the answers agree tolerably well with results found in other surveys, they may insofar be trusted. If, for instance, children report that they sleep on the average nine hours a day, and these results are similar to the findings of other studies, they may be trusted.

(b) The results should vary with varying conditions. Answers to the question, "Are you up at seven o'clock in the morning?" show that nearly 2 per cent more answer *yes* in May than in October, as would be expected, because the days are longer and the mornings lighter in May than in October.

(c) When conditions are constant, the answers also should be constant. To the question, "Is there a bathtub in your home?" the answers for grades 4, 5, 6, 7, and 8 were 86 per cent, 86 per cent, 87 per cent, 88 per cent, 89 per cent. This consistency is to be expected, as there is no factor which operates to make the

percentages of homes possessing bathtubs vary from grade to grade.

(d) Testimony of individual schools or groups of schools. In a school where more than 95 per cent of the pupils were either Japanese or Chinese, the percentages of tea-drinkers was 54, while the percentage for the whole city was only 19. In one seventh grade in a school in the best residential district, 100 per cent of the pupils lived in homes which had bathtubs, while in an entire seventh grade in an outlying district the percentage was 28. This is the kind of consistency in the answers for groups which tends to prove that the returns possess a high degree of accuracy.

In psychological questionnaires (for example, the Woodworth Psychoneurotic Inventory) the matter of veracity does not enter so strongly as a disturbing factor. These questionnaires are given under standardized conditions, and their significance is determined by their correlations with other factors. If certain pupils do not choose to answer the questions truthfully, this fact is taken care of by the correlations. It is not the question whether the pupil will tell the truth or not, but how his answers compare with those of some one else when asked under standardized conditions, and the way these differences are related to other factors.

As measures of attitudes, questionnaires have been under fire in this matter of veracity. There may oftentimes be a tendency to answer for effect. We have referred (pp. 150, 151) to the studies of Persing and Hartshorne and May to show that under certain conditions there is divergence between what pupils say and what they do. The common impression is that a person will act to further his own interests, and when asked a question, will reply in order to produce the most favorable impression. However, the writer has long been of the impression from casual observation that pupils in school will answer these questionnaires rather truthfully. One reason for this is that pupils confuse these questionnaires with tests, and with the general school attitude of trying to do their best on a test they seriously try to give truthful answers on questionnaires concerning even their most personal affairs. Another reason is that the same motive that causes a person to make a good impression on a questionnaire would also make him wish to give a good impression in conduct. In his book *The Nature of Conduct* (pp. 231-233) the author expresses himself as follows on this point:

"Verbal expression does not necessarily truthfully portray inner facilitation or resistance. In cases of dissimulation one may really speak contrary to actual readiness. Boasting is an example; rationalization is another. A child may claim to be unafraid, for instance, of going on a boat or into a dark room but may balk at the actual act. Inexperienced persons, on this account, distrust those questionnaires which have recently been devised to measure a person's interests or attitudes with special reference to his progressiveness or conservativeness. The objection raised is that it is easy to state one's beliefs or interests one way or the other regardless of one's real interests or beliefs. The answers may be guided by one's opinion as to how other people will judge one by them. It happens, however, that in actual practice such dissimulation takes place only infrequently. One's natural (although probably learned) reaction is to make the verbal statement harmonious with the actual readiness or unreadiness. It would also seem that if a person is influenced by what other people say or think, he is usually influenced not only in his words but also in his internal readinesses and unreadinesses. Most persons have learned to make their words and thoughts match their actions. It is the exception for a teetotaler to say that he does not believe in prohibition. Most men who say they disbelieve in prohibition will suit action to the word on the proper occasion. A poll or straw ballot on a group's attitude toward prohibition is apt to give a pretty accurate picture not only of what the group professes but of their actual conduct. It is for this reason that these questionnaires designed to tap attitudes have proved themselves rather reliable and significant indicators of conduct trends."*

Recent experimental data published by the Character Education Inquiry tend to substantiate the above statements. Quotation will be made only from the conclusions as stated by May and Hartshorne, as a description of their experimental technique would be somewhat involved:

"Over against the fact that 89 per cent (of children) stated that it was their duty to read the Bible every day (apparently a conventional response) must be set the fact that very few of these children changed their answers when confronted with the answer sheet which showed that the 'standard' did not regard it as their duty to do so. . . . We may say, on the one hand, that the moral knowledge scores, therefore, which differ widely from child to child, are not merely efforts to repeat the school standards, but represent something more fundamental. Or, on the other hand, we may say that the tendency to make a good ap-

* Symonds, P. M., *The Nature of Conduct* (The Macmillan Company, 1928). By permission of The Macmillan Company, publishers.

156 Diagnosing Personality and Conduct

pearance does not correlate with moral knowledge. . . . The consistency of results, as found by correlating comparable forms of the same test, indicates that what is stated as moral knowledge has a certain coherence and stability, and whether lived up to or not, points the way to action that is regarded as proper."*

With reference to the relation of verbal attitude and conduct these authors say:

"Ninety-one per cent of those who say it is all right to let another pupil copy your work and hand it in as actually his own cheated themselves. One hundred per cent of those preferring to smash the slot machine to recover their lost nickel actually cheated on a test. Ninety-three per cent of those who thought it right for John to cheat in order to help his class win actually cheated themselves. It is noteworthy that these high agreements among the cheaters are in regard to cheating in two cases, to property in the third, and *not in any instance to other types of behavior*. This is somewhat surprising, since one would not expect a cheater to wear his heart on his sleeve."**

The present writer's own interpretation of the data given in the article would not be so strongly in favor of the position quoted. For example, there were many more pupils who said it was wrong to cheat and did cheat than said it was right to cheat and did cheat. The numbers on which the percentages are based are very small. The zero correlation which these investigators found between moral knowledge and conduct proves nothing, except that both knowledge and conduct are very specific.

The safest plan is to wait for adequate experimental results before deciding as to the dependence one can place on the truthfulness of answers to a questionnaire. Eventually we shall know under what conditions the truth can be expected and when it cannot be expected.

Methods of Scoring Questionnaires

In the fact-finding questionnaire there is no occasion for summing the answers. The answer to each question conveys an important item of information to be tabulated separately. In the

*Hartshorne, H., and May, M. A., "Testing the Knowledge of Right and Wrong," *Religious Education*, 21: 419 f. (August, 1926).

**Hartshorne, H., and May, M. A., in *Religious Education*, 21: 627 (Dec., 1926).

psychological questionnaire, where all of the questions attempt to get reactions on various phases of human attitude or interest, it is customary to sum the answers. For instance, if the questions are designed to show the subject's reaction to various social issues, it is possible to sum the responses so as to show a radical or conventional tendency. In the interest questionnaires one can sum the answers so as to get a measure of interest in broadly designated activities or occupations. Certain investigators, to be sure, have refused to sum their answers. Reasons given are that each question is valuable on its own account, that to give a single score would obscure differences in the way in which individual questions were answered, and that a total score would have no meaning, since it stands for the answers to a miscellaneous group of questions. Much depends on the use to be made of the results of the questionnaire. If it is to become a basis of group discussion, or to reveal to an individual his own idiosyncrasies, or to guide a doctor in giving advice or administering treatment, then the answer to each question is of value. But if it is to be used to measure attitude in some general sphere such as racial relations, religion, or the economic order, or to measure interest in some phase of life, or to measure the degree to which successful adaptation has been made to some part of the environment, then there is every reason for obtaining a total score. It is a satisfaction to report that sufficiently high correlations have been found between the halves of questionnaires designed to measure such characteristics as psychoneurotic adjustment, introversion-extroversion, inferiority complex, fair-mindedness, studiousness, and ascendancy-submission to give us firm ground for believing that such qualities are real characteristics of the personality and that they may be measured by the questionnaire-rating scale method.

The method of scoring a questionnaire, once it has been decided that scoring is advisable, is largely determined by considerations similar to those described in selecting the items. One may use a common-sense scoring method which gives credit only for answers which apparently are in the direction of or show harmony with the contact or characteristic being measured. Such a key may be derived most simply from a *prima facie* consideration of the questions. Better than this is to obtain the direction of the questions by a consensus of judgment of several competent persons. The present writer, for instance, in his "Social Attitudes

158 Diagnosing Personality and Conduct

Questionnaire"* determined the liberal side of questions on social issues by asking five experts to answer the questions with what they considered to be the liberal position. Items on which there was no practical agreement were discarded.

Another scoring method consists in trying out the questions first and determining the key from the results. This may be accomplished in two ways. The first of these is to tabulate the responses on two sharply differentiated groups already separated either by social selection or by ratings on the quality or characteristic in question. The other method is first to score the questionnaire, using a common-sense or *a priori* key, and then to check the answers to each item against high or low total scores on the questionnaire. This can be done in various ways, of which an elaborate one is to compute the bi-serial r for the answers to each item against the distribution of scores. A simpler method is to compute the number of *yes* answers for sub-groups making above-average and below-average scores in the group originally measured. In this case the key should give credit for the answer *yes* if a larger percentage of high-scoring than low-scoring persons answers it that way. Where the group is large, one may abridge the work by using the upper and lower 25 or 20 per cent of the group.**

In case more than a twofold or dichotomous response (e.g., *Yes*, *?*, *No*) is obtained on a question, the problem arises of assigning credit for different answers. This problem becomes acute when a graphic rating scale is used for entering the answer, as in the "Colgate Personal Inventory." Here one is confronted with the task of evaluating the position of the rating for the purpose of using it most effectively in determining a total score on the questionnaire. Laird's † solution was to determine a point on each rating scale above which lay 25 per cent of the answers in the direction of introversion and then merely to give one point credit for any answer which lay inside this extreme area. This method,

* *Journal of Educational Psychology*, 6:316-322 (1925).

** Kelley has determined that the optimum differentiation is obtained when answers made by the 27% of the group making the highest scores are compared with the answers made by the 27% of the group making the lowest scores. See Jensen, M. B., *Objective Differentiation Between Three Groups in Education*, Genetic Psychology Monographs, vol. 3, no. 5, 1928.

† Laird, D. A., "Detecting Abnormal Behavior," *Journal of Abnormal and Social Psychology*, 20: 128-141 (1925).

however, seems to the present writer to throw away most of the rewards to be had for the trouble of using a graphic scale. Some such simple device as giving one, two, three, four, or five credits according to the fifth of the scale in which an answer lies seems an easy way of making more effective use of the sharper differentiation which the graphic scale affords. Freyd* met the same problem when he had to construct a scoring scheme for answers to his interest questionnaire, which called for encircling L!, L, ?, D, or D! His original decision to give credit for only one of the symbols encircled in each item was not followed by subsequent workers. Exactly what methods strike the best balance between making the most effective use of the fineness of discrimination which the self-ratings afford and avoiding unnecessary difficulty in scoring is a task for further research to decide. If Laird's solution of merely giving a credit for answers in certain areas of each scale is used, the method of answering the questions could surely be much simplified without becoming less valuable. Likewise if only one or two of the answers is to be scored when a five-choice alternative is offered, the questionnaire would probably yield almost as satisfactory results if it offered less than five alternatives. In general we should offer no more possible alternatives as answers than it is decided to use in scoring.

This leads us to the matter of weighing items. It seems only common sense that answers to the more valid items should have greater significance and perhaps should be given more weight than answers to less valid items. Two methods for doing this have been suggested. One developed by Kelley and described by Cowdery** is to be used where there are dichotomous answers. The weight to be used is

$$b = \frac{\phi}{(1 - \phi^2)\sigma}$$

where

$$\phi = \frac{ad - bc}{\sqrt{(a+c)(b+d)(a+b)(c+d)}}$$

where a and b are the numbers answering *yes* and *no* to the question; and c and d are the numbers in each of the two groups

* Freyd, M., "The Personalities of the Socially and Mechanically Minded," *Psychological Monographs*, 33, No. 4 (1924).

** Cowdery, K. M., "Measurement of Professional Attitudes," *Journal of Personnel Research*, 5:131-141 (1926-1927).

between which the questionnaire is expected to differentiate. Strong discarded this method in his work because he found the results correlated almost perfectly (except for extreme values) with those to be had by merely taking differences between percentages of subjects answering the questions correctly. Differences between percentages were then coded down into a set of weights to be used in scoring.

Watson* has used an adaptation of the McCall-Long method of weighting scores. He describes the steps of his plan as follows:

"1. Build the criterion.

"2. List across the top of a sheet every possible response to every element in the test.

"3. Tabulate the answers for each person in the criterion group, and enter his criterion score in the column corresponding to every answer which he has made. Thus, suppose we had an individual whose criterion score in the test questions was 95. Suppose this person answered true to the first question, true to the second, checked the second alternative in the multiple-choice question, and ranked the ranking questions in the order 5, 3, 2, 1, 4. Then 95, the individual score would be listed in every column corresponding to the response he made. The same would be done with the total scores made by other individuals.

"4. Add each column and divide by the number of entries in the column. This will give the average value of each answer in terms of criterion score. At the same time add the column of criterion scores so that we may find the average of the criterion scores.

"5. Subtract the average criterion score in algebraic fashion from the average of each column. The resulting figure will give, if positive, the amount of credit that ought to be given for that answer; if negative the amount which should be subtracted from the score for any person who makes the particular response."

Watson adds the following comment,

"If the test be scored in this fashion it will give high scores to the kind of people who stood high in the criterion, and it will give low scores to the kind of people who stood low in the criterion. Any elements which were answered in the same fashion by everyone receive zero rating. Any answers which were answered by nobody are eliminated. Any elements which were answered equally often by people who stood high in the scale and people who stood low in the scale receive zero value. The difference between

* Watson, G. B., *Experimentation and Measurement in Religious Education* (Association Press, 1927), p. 155 f.

the amount of positive credit which can be given for the right answer to a test question, and the amount of penalty which is given for the poorest answer, is a rough indication of the value of that element."

Two tests should be applied to any system of weighting items before the scheme is adopted. One test is whether the weights are consistently retained when derived from a new group. Unless the weights hold in a new situation, they should not be used for scoring. The other test is whether the correlation with a criterion using the weights is sufficiently superior to the correlation without using the weights. If the difference in correlations is insignificant, there can be no advantage in using the more complicated scoring device, because, as a general principle, a scoring scheme should be as simple as possible when it may be hoped that the test will be popularly used.

Scores of amount and of direction. It is possible to score certain types of questionnaire in two or more ways so as to bring out different characteristics of the response. Where the subject answering the questions is instructed to mark only those items about which he has definite opinions or feelings, the number of responses made has a significance. This type of score is used by Pressey* with his X-O tests—Pressey calls it an "affectivity" score. Watson** takes the number of extreme answers as a measure of prejudice in his tests of fair-mindedness. Besides this total score, it is possible to take into account the direction of the score by considering the direction of the answer to certain groups of items. This type of score Pressey calls an "idiosyncrasy" score and by means of it he measures such abnormal responses as disgust, fear, sex, and self-feeling. Watson also uses a differential score to measure direction of prejudice in such fields as the economic order, race problems, and religious issues.

Reliability of the Questionnaire

The problem of accuracy of measurement in the physical sciences come down to the relatively simple matter of noting the

* Pressey, S. L., and Chambers, O. R., "First Revision of a Group Scale for Investigating the Emotions, with Tentative Norms," *Journal of Applied Psychology*, 4: 97-104 (1920).

** Watson, G. B., *The Measurement of Fairmindedness*, Teachers College Contributions to Education, No. 176, pp. 13-15.

variations resulting from repetitions of a measurement. In psychological testing, however, it is impossible to obtain an exact repetition because, once a person has taken a test, his familiarity with it precludes setting up exactly the same conditions. The problem is solved, however, by recognizing that any test represents but a sampling of the field tested, and the repetition is accomplished by testing with another sample. The correlation between the scores on these two supposedly equivalent tests given at different sittings is used as a measure of reliability and is called the reliability coefficient.

In determining the reliability of questionnaires it is often difficult to gather another set of questions which would represent a sampling of the field equivalent to the original one. A method of circumventing this difficulty is to correlate the scores on randomly chosen halves of the questionnaire, afterwards correcting this reliability coefficient of one half against the other half by the Brown-Spearman formula, which enables one to estimate the correlation of the whole questionnaire against another similar one. This method does not take into account differences in the subject between two sittings. It is really only a measure of the internal consistency of the questions, a very important consideration, however, when one is measuring in a new field. The method of repeating the questionnaire has sometimes been employed, but no one can know to what degree the questions are remembered and deliberate changes made in the answers. Cady,* following Kelley's suggestion, has used the ingenious device of rewriting the questionnaire so that the answers to all questions are reversed. For instance

(First form) Do you ever have a strong desire to set fire to something?

(Second form) Do you dislike the idea of setting fire to something to see it burn?

Although on the surface of things this seems to ensure comparability of the two forms, it really produces a new set of questions. Concerning this Cady states,

"A lack of clearness and directness in many of the questions will be apparent. It is very difficult to change a statement from

* Cady, V. M., *The Estimation of Juvenile Incurability*, Journal of Delinquency Monographs, November 2, 1923.

the natural direct way of expression into a question that says the same thing, yet requires an opposite reply. The result is a number of expressions which are practically double negatives. If they are not of this nature, they often do not mean quite the same thing, though perhaps near enough for practical purposes. A portion of these difficulties was overcome by explanations made to the subjects which helped to clear things up for those of less intelligence and yet did not introduce new meanings. . . . It is open to question, however, whether this device of a reversed question would be as valuable as, say, 100 carefully selected questions divided into two forms and given at different periods."

A few typical reliabilities of questionnaires taken from the experimental literature are as follows:

TABLE 22
RELIABILITY COEFFICIENTS OF QUESTIONNAIRES

<i>Test</i>	<i>Reported by</i>	<i>Group tested</i>	<i>Number of items</i>	<i>Method of determining reliability</i>	<i>Coefficient of reliability</i>
Woodworth Psychoneurotic Inventory	Mathews	drafted men	116	Split halves	.90
Woodworth-Mathews Questionnaire	Mathews	280 boys 12, 13, 14 years old	75	Split halves	.667
Woodworth-Mathews Questionnaire	Mathews	26 boys aged 9-19	75	Retest	.369
Woodworth-Mathews Questionnaire	Mathews	29 girls aged 9-19	75	Retest	.697
Woodworth-Cady Questionnaire	Cady	Boys 13 and 14 years old	60	Correlation with duplicate form	.55
Woodworth-Cady Questionnaire	Cady	Boys 13 and 14 years old	About 30	Correlation with duplicate form	.49
Woodworth-Cady Questionnaire	Cady	150 boys 13 and 14 years old	About 30	Correlation with duplicate form	.47
Woodworth-House Questionnaire	House	68 West Point students	100	Retest	.714
Woodworth-House Questionnaire	House	58 Harvard students	99	Retest	.845
Colgate Mental Hygiene Test	Hoitsma	College students	75	Retest	.85
Colgate Mental Hygiene Test	Hoitsma	College students	75	Split halves	.79
Pressey X-O					
Affectivity Scores, test 1	McGeoch & Whiteley	College students	125	Retest	.85
test 2	McGeoch & Whiteley	College students	125	Retest	.86

test 3	McGeoch & Whitely	College students	125	Retest	.82
test 4	McGeoch & Whitely	College students	125	Retest	.87
Idiosyncrasy Scores, test 1	McGeoch & Whitely	College students	25	Retest	.70
test 2	McGeoch & Whitely	College students	25	Retest	.55
test 3	McGeoch & Whitely	College students	25	Retest	.77
test 4	McGeoch & Whitely	College students	25	Retest	.43
Colgate Personal Inventory, C-1	Hoitsma	College students	53	Retest	.674
Colgate Personal Inventory, C-1	Hoitsma	88 college students	53	Split halves	.45
Colgate Personal Inventory, C-3	Laird	College students	41	Two associates	.85
Colgate Personal Inventory, C-3	Laird	College students	41	Retest—self-ratings	.90
Colgate Personal Inventory, C-1	Conklin	164 college students	53	Split halves corrected by Spearman-Brown	.72
Colgate Personal Inventory, C-2	Guthrie	327 college students, Univ. of Washington		Split halves corrected by Spearman-Brown	.60
Heidbreder Introversion-Extroversion	Heidbreder	College students	54	Self-associates ratings	.55
Heidbreder Introversion-Extroversion	Heidbreder	College students	54	Associates—Associates ratings	.40
Marston Introversion-Extroversion	Marston	Children 2-6 years of age	26	Split halves	.89

TABLE 22—Continued

<i>Test</i>	<i>Reported by</i>	<i>Group tested</i>	<i>Number of items</i>	<i>Method of determining reliability</i>	<i>Coefficient of reliability</i>
Marston Introversion-Extroversion	Marston	Children 2-6 years of age	26	Ratings of two judges	.71
Conklin Introversion-Extroversion	Conklin	352 college students	40	Split halves corrected by Spearman-Brown	.95
Conklin Introversion-Extroversion	Conklin	134 college students	40	Split halves corrected by Spearman-Brown	.92
Cowdery Interest Report Blank	Cowdery	Experienced men	264	Split halves corrected by Spearman-Brown	.827
Cowdery Interest Report Blank	Cowdery	Graduate students	264	Split halves corrected by Spearman-Brown	.818
Watson Survey of Public Opinion on Some Religious and Economic Issues	Watson	Not stated	489	Split-halves corrected by Spearman-Brown	.96
Symonds Studiousness Questionnaire	Symonds	80 high school seniors	100	Split halves corrected by Spearman-Brown	.846
Symonds Studiousness Questionnaire	Symonds	213 high school seniors	100	Split halves corrected by Spearman-Brown	.820
Symonds Social Attitudes Questionnaire	Symonds	102 college freshmen	115	Duplicate form	.67

Zeleny True-False Opinion Test	Zeleny	35 college students	68	Correlation of two parts corrected by Spearman-Brown	.893
Zeleny True-False Opinion Test	Zeleny	58 college students	68	Correlation of two parts corrected by Spearman-Brown	.894
Harper Test of Social Attitudes	Harper	170 graduate students in education	71	Split halves	.782
Harper Test of Social Attitudes	Harper	Rural and village teachers	71	Split halves	.751
Harper Test of Social Attitudes	Harper	Undergraduates in city training school	71	Split halves	.817
Jones Opinion Test	Jones	College students	25	Split halves	.71

168 Diagnosing Personality and Conduct

These reliability figures compare very favorably with other similar reliabilities obtained from recall multiple-response and true-false tests. Ruch * gives the following table which summarizes work on reliability with objective tests.

TABLE 23
RELIABILITY OF TYPICAL OBJECTIVE SUBJECT-MATTER TESTS
(from Ruch)

RELIABILITY OF 100 ITEMS					
	Recall	Five- response	Three- response	Two- response	True- false
Toops764	.715673
Ruch-Stoddard896	.886	.748	.849	.714
Ruch-Stoddard Social Study...	.950	.882	.890	.843	.837

Probably the same factors influence the reliability of questionnaires that influence the reliability of tests, with, however, certain factors assuming particular importance in determining the reliability of the questionnaire. It has been found that questionnaires are more reliable for adults than for children. There are three factors which may go toward explaining this. In the first place, since adults are probably better able to observe and introspect than are children, their answers to questions should be more accurate and consistent, as is observable, for example, in the Woodworth Psychoneurotic Inventory. In the second place, it is probably true that most personal experiences, characteristics, etc., are better developed in adults than in children, and hence stand in a clearer light and are not so difficult to distinguish. This phenomenon would apply particularly to interest questionnaires. A third factor is the matter of intelligence necessary to understand the items. For younger children, many of the questionnaires become intelligence tests or vocabulary tests because of the inherent difficulty of the concepts and the language employed in them. Children would have difficulty in answering the following lines taken at random from the Pressey X-O Tests:

extravagance	sportiness	boasting	deformity	talking-back
fire	nervousness	germs	insult	disfigurement
Mowgli	Tarzan	D'Artagnan	Hamlet	Gallahad

* Ruch, G. M., and others, *Objective Examination Methods in the Social Studies* (Scott, Foresman and Company, 1926), p. 83.

In cases where the issue becomes one of not understanding the item, the score tends to reduce to 50 per cent of the items answered, and chance becomes the supreme factor in determining the score.

Norms

Shall norms be given for questionnaires? By all means, since norms for typical groups which have answered the questionnaire help to interpret the results of measurement. Only by considering the present case in the light of previous experience can judgment be made as to the significance of the results. Besides the average score made by the group, percentile scores should be given, so that one may know what percentage of the individuals in the group tested fall above a certain score. In many cases norms or percentile scores for contrasted groups are necessary so that one may get at the typicality of a given score for a given group. For instance, norms should be available for the Woodworth Psychoneurotic Questionnaire in the case of normal groups, psychoneurotic groups, and delinquent groups, etc., thus permitting estimates of the probability that any score will fall in any of those groups. Likewise norms and distributions for the Cowdery Vocational Interest Blank should be provided for each group to be differentiated. Age norms seem of less value in the case of questionnaires than in the case of tests. The reason for this is that in the former it is difficult to perceive a real growth or change due to age. Whatever changes do occur during the ages of growth can probably be explained as due to an increase in ability to read or to greater comprehension of the situations described.

We should be on our guard against making a false interpretation of the meaning of norms. A common error is to accept norms as standards or goals, an error quite likely to be made with the results of attitudes questionnaires. An individual tends to feel a certain degree of satisfaction if he reaches the average attitude of his group. Though to think or believe like the group we respect is all that we usually desire, the group *may* not be right. Perhaps we ought to be more open-minded or more tolerant than any group measured. Science can never tell us what we ought to do merely from studying that which has been done. However, in reaching a decision as to what should be done, a knowledge of the existing state of affairs is helpful.

Types of Distributions Yielded by Questionnaires

Most questionnaires yield distributions which are practically normal. It is significant that questionnaires designed to show characteristics of adjustment such as psychoneurosis, introversion-extroversion, superiority-inferiority, and ascendancy-submission all show approximately normal distributions. Most persons, then, are not introvert or extrovert—they are ambivert. Most persons strike a balance between feelings of superiority and inferiority or between reactions typical of ascendancy and submission. These dichotomous or twofold classifications, which are so common in our everyday interpretations and generalizations, are not borne out by the test results. The average, however, may not be exactly at the dividing point. Heibredner found that the normal person tended to answer more of her questions like the extroverts than like the introverts. It is more unusual to be introverted than extroverted. Furthermore, even though a distribution is normal for other characteristic differences, and few people are extreme, one type of adjustment is usually rarer than the other.

Even in our attitudes and interests we tend to show a normal distribution. Few boys are extremely studious, and few are extremely the opposite—the majority are tolerably studious. Most persons strike a balance between the radical and the conservative. We are all radical on some issues and conservative on others. Since for most persons these tendencies balance up, it is the outstanding or peculiar person only who is marked as the ultra-radical or the ultra-conservative.

One exception to this was discovered by Allport* who studied attitudes by asking students to select from a submitted list the statement regarding an issue which was most acceptable to them. He found that on some issues—prohibition, for example—there was a definite cleavage of opinion, with few taking a middle ground, such that they could be called neither wet nor dry in sentiment.

But Thurstone,** using the same method, believes that this

* Allport, F. H., and Hartman, D. A., "Measurement and Motivation of a Typical Opinion in a Certain Group," *American Political Science Review*, 19: 735-760 (1925).

** Thurstone, L. L., "Attitudes Can Be Measured," *American Journal of Sociology*, 33: 529-554 (Jan., 1928).

characteristic in Allport's results was due to his scale. It was thick with items in the middle and thin with items on the ends. That is, Thurstone thinks, the form of the distribution of attitudes is determined by the scale through the selection of items included in it. There is just a touch of inconclusiveness in the argument here. If we start with the assumption that the distribution should be normal, and then make up the scale to fit the assumption, naturally the results of using it will show a normal distribution. This problem is best answered by considering the distribution formed by submitting a long list of items chosen by random sampling. Such distributions are usually normal.

Sex differences. Most questionnaires, unlike tests, show distinct sex differences. Women, for instance, have been found by Laird* to be more introverted than men. Symonds** found women students more conservative than men students. Indeed, the differences are so apparent in interest questionnaires that separate sets of questions must be assembled to measure adequately the interests of the two sexes—a finding markedly out of line with the results of tests, which seldom show clearly marked sex differences.

Racial differences. The Woodworth Psychoneurotic Inventory has yielded distinct racial differences.† Investigators, however, incline to the opinion that these differences, instead of indicating fundamental trends in racial constitution, are really an expression of varying mores, customs, and habits of thought and action, a conclusion which, if correct, would imply that the results of any questionnaire of adjustment, interests, or attitudes must be interpreted in the light of an individual's or a group's background. Differences disclosed by questionnaires probably reflect the effect of environment, particularly the social environment, rather than some inborn native tendency of the personality.

REFERENCES

1. ALEXANDER, C., *School Statistics and Publicity* (Silver, Burdett and Company, 1919).
2. ALEXANDER, C., and others, *Educational Research* (Bureau

* Laird, D. A., "How Personalities Are Found in Industry," *Industrial Psychology*, 1: 654-662 (1926).

** Symonds, P. M., "A Social Attitude Questionnaire," *Journal of Educational Psychology*, 16: 316-322 (1925).

† Abel, J. M., "National Differences and Preferences in Beliefs," *Journal of Comparative Psychology*, 6: 21-41 (1926).

- of Publications, Teachers College, Columbia University, 1927), 42 pp.
3. BAILEY, W. B., and CUMMINGS, J., *Statistics* (A. C. McClurg & Co., 1917), chs. I-III.
 4. BOWLEY, A. L., *Elements of Statistics* (New York: Charles Scribner's Sons; London: P. S. King & Co., 1907, 1926), 462 pp., pp. 33-62.
 5. BURK, F., "On a Certain Questionnaire," *School and Society*, 15: 170-173 (Feb. 11, 1922).
 6. BUCKINGHAM, B. R., "Questionnaires," *Journal of Educational Research*, 14: 54-58 (June, 1926).
 7. BUTTERFIELD, E. W., "Plenary Inspiration of the Dotted Line," *Educational Review*, 69: 1-4 (Jan., 1925).
 8. ———, "Professional Ethics and the Questionnaire," *School and Society*, 11: 55-56 (Jan., 1920).
 9. CHADDOCK, R. E., *Principles and Methods of Statistics* (Houghton Mifflin Company, 1925), pp. 375-378, 388-391.
 10. DAVIS, W. W., "Questionnaire Method in Health Education," *Elementary School Journal*, 23: 373-386 (Jan., 1923).
 11. Editorial, "Job Analysis and After," *Journal of Educational Method*, 5: 323-324 (April, 1926).
 12. EDWARDS, A. L., "Tools for Building the Curriculum, Surveys and Questionnaires," *Journal of Home Economics*, 17: 194-198 (April, 1925).
 13. FAIR, G. M., "Questionnaire," *Public Health Nurse*, 15: 229-230 (May, 1923).
 14. FRANKEN, R. B., "Formulating Questionnaires," *Advertising Fortnightly*, Jan. 30, 1924.
 15. GOOD, C. V., *How to Do Research in Education* (Warwick & York, 1928).
 16. HARTSHORNE, H., and MAY, M. A., *Studies in Deceit*, Studies in the Nature of Character, Vol. I (The Macmillan Company, 1928), 306 pp.
 17. HEINEMAN, T., "Diploma of the Question Mark," *Advertising Fortnightly*, July 18, 1923.
 18. HOBSON, A., "The Use of the Correspondence Method in Original Research," *Quarterly Publication of the American Statistical Association*, 15: 210-218.
 19. KELLEY, T. L., *Scientific Method—Its Function on Research in Education* (The Ohio State University Press, 1929), Lecture II.
 20. KING, W. I., *Elements of Statistical Method* (The Macmillan Company, 1916), chs. V-VI.
 21. KOOS, L. V., *The Questionnaire in Education* (The Macmillan Company, 1928).
 22. LINDSAY, E. E., "Questionnaires and Follow-Up Letters," *Pedagogical Seminary*, 28: 303-307 (Sept., 1921).

23. LINK, H. C., *Employment Psychology* (The Macmillan Company, 1919).
24. MCCALL, W. A., *How to Experiment in Education* (The Macmillan Company, 1923).
25. National Committee on Research in Secondary Education, "An Outline of Methods of Research with Suggestions for High School Principals and Teachers," U. S. Bureau of Education Bulletin, No. 24 (1926).
26. National Education Association Research Bulletin, "The Questionnaire," 8: No. 1 (Jan., 1930).
27. PERRY, H. E., "The Questionnaire Method," *Journal of Applied Sociology*, 10: 155-158 (1925-1926).
28. PERSING, K. M., "Morals and Chemistry," *Educational Review*, 72: 164-168 (Oct., 1926).
29. POFFENBERGER, A. T., *Psychology of Advertising* (A. W. Shaw Company, 1925), 632 pp.
30. "Questionnaires," *School Review*, 34: 649-50 (Nov., 1926).
31. REEDER, W. G., *How to Write a Thesis* (Public School Publishing Company, 1925).
32. RUGG, H. O., *Statistical Methods Applied to Education* (Houghton Mifflin Company, 1917).
33. SCHLUTER, W. C., *How to Do Research Work* (Prentice-Hall, Inc., 1926), 138 pp.
34. SECRIST, H., *Introduction to Statistical Methods* (The Macmillan Company, 1925).
35. TOOPS, H. A., "Validating the Questionnaire Method," *Journal of Personnel Research*, 2: 153-169 (Aug.-Sept., 1923).
36. ———, "Returns to Follow-Up Letters to Questionnaires," *Journal of Applied Psychology*, 10: 92-101 (March, 1926).
37. TROW, W. C., *Scientific Method in Education* (Houghton Mifflin Company, 1925).
38. WELLS, F. L., "Report on a Questionnaire Study of Personality Traits with a College Graduate Group," *Mental Hygiene*, 9: 113-127 (1925).
39. WHITNEY, F. P., "Questionnaire Craze," *Educational Review*, 68: 139-140 (Oct., 1924).
40. WADMORE, M., "An Inquiry into Children's Ideas in Social and Industrial Questions," *Journal of Experimental Pedagogy*, 6: 303-317 (1922).
41. WYLIE, A. T., "To What Extent May We Rely upon the Answers to a School Questionnaire?" *The Journal of Educational Method*, 6: 252-257 (Feb., 1927).

Chapter V

ADJUSTMENT QUESTIONNAIRES

The Psychoneurotic Inventory and Its Derivatives

A FORM of questionnaire for measuring psychoneurotic tendencies which has enjoyed extensive use, had its inception in 1917 during the World War. At that time a method was needed for diagnosing the ability of men to adjust themselves satisfactorily to the stresses and strains of military life. Since it was obvious that the psychologists who were available could not possibly interview personally every drafted man, a paper-and-pencil questionnaire was drawn up by Woodworth, then chairman of the Committee on Emotional Fitness appointed by the National Research Council. Woodworth studied the symptoms of men who had difficulty in adjusting themselves to trying conditions, using such sources as McCurdy's *War Neuroses*. Over 200 questions were assembled, each to be answered by *yes* or *no*. These were tried on Columbia College students and drafted men. Questions for which the percentage of "unfavorable" replies was large were either omitted or "stiffened," so that as a result of this preliminary experience the list was reduced to 116. Although the armistice came before this final list could be extensively studied, a report of its use in army hospitals is contained in H. L. Hollingworth's *Psychology of the Functional Neuroses*. We shall present Woodworth's list of 116 questions, as it forms the basis of subsequent work. The unfavorable answers are italicized. The score is the number of unfavorable responses. This questionnaire, originally dubbed "Personal Data Sheet" in order to allay suspicions as to its real purpose, has been called in psychological literature the "Woodworth Psychoneurotic Inventory."* (See 11, pp. 171-176.)

*Published by C. H. Stoelting Company. From Franz, *Handbook of Examination Methods*. By permission of The Macmillan Company, publishers.

Woodworth Psychoneurotic Inventory

- | | | |
|--|-----|----|
| 1. Do you usually feel well and strong? | yes | no |
| 2. Do you usually sleep well? | yes | no |
| 3. Are you frightened in the middle of the night? | yes | no |
| 4. Are you troubled with dreams about your work? | yes | no |
| 5. Do you have nightmares? | yes | no |
| 6. Do you have too many sexual dreams? | yes | no |
| 7. Do you ever walk in your sleep? | yes | no |
| 8. Do you ever have the sensation of falling when going to sleep? | yes | no |
| 9. Does your heart ever thump in your ears so that you cannot sleep? | yes | no |
| 10. Do ideas run through your head so that you cannot sleep? | yes | no |
| 11. Do you feel well rested in the morning? | yes | no |
| 12. Do your eyes often pain you? | yes | no |
| 13. Do things ever seem to swim or get misty before your eyes? | yes | no |
| 14. Do you often have the feeling of suffocating? | yes | no |
| 15. Do you have continual itching in the face? | yes | no |
| 16. Are you bothered much by blushing? | yes | no |
| 17. Are you bothered by fluttering of the heart? | yes | no |
| 18. Do you feel tired most of the time? | yes | no |
| 19. Have you ever had fits of dizziness? | yes | no |
| 20. Do you have queer, unpleasant feelings in any part of the body? | yes | no |
| 21. Do you ever feel an awful pressure in or about the head? | yes | no |
| 22. Do you often have bad pains in any part of the body? | yes | no |
| 23. Do you have a great many bad headaches? | yes | no |
| 24. Is your head apt to ache on one side? | yes | no |
| 25. Have you <i>ever</i> fainted away? | yes | no |
| 26. Have you <i>often</i> fainted away? | yes | no |
| 27. Have you ever been blind, half-blind, deaf, or dumb for a time? | yes | no |
| 28. Have you ever had an arm or leg paralyzed? | yes | no |
| 29. Have you ever lost your memory for a time? | yes | no |
| 30. Did you have a happy childhood? | yes | no |
| 31. Were you happy when 14 to 18 years old? | yes | no |
| 32. Were you considered a bad boy? | yes | no |
| 33. As a child did you like to play alone better than to play with other children? | yes | no |
| 34. Did the other children let you play with them? | yes | no |
| 35. Were you shy with other boys? | yes | no |
| 36. Did you ever run away from home? | yes | no |

176 Diagnosing Personality and Conduct

37. Did you ever have a strong desire to run away from home? yes no
38. Has your family always treated you right? yes no
39. Did the teachers in school generally treat you right? yes no
40. Have your employers generally treated you right? yes no
41. Do you know of anybody who is trying to do you harm? yes no
42. Do people find fault with you more than you deserve? yes no
43. Do you make friends easily? yes no
44. Did you ever make love to a girl? yes no
45. Do you get used to new places quickly? yes no
46. Do you find your way about easily? yes no
47. Does liquor make you quarrelsome? yes no
48. Do you think drinking has hurt you? yes no
49. Do you think tobacco has hurt you? yes no
50. Do you think you have hurt yourself by going too much with women? yes no
51. Have you hurt yourself by masturbation (self-abuse)? yes no
52. Did you ever think you had lost your manhood? yes no
53. Have you ever had any great mental shock? yes no
54. Have you ever seen a vision? yes no
55. Did you ever have the habit of taking any form of "dope"? yes no
56. Do you have trouble in walking in the dark? yes no
57. Have you ever felt as if some one was hypnotizing you and making you act against your will? yes no
58. Are you ever bothered by the feeling that people are reading your thoughts? yes no
59. Do you ever have a queer feeling as if you were not your old self? yes no
60. Are you ever bothered by a feeling that things are not real? yes no
61. Are you troubled with the idea that people are watching you on the street? yes no
62. Are you troubled with the fear of being crushed in a crowd? yes no
63. Does it make you uneasy to cross a bridge over a river? yes no
64. Does it make you uneasy to go into a tunnel or subway? yes no
65. Does it make you uneasy to cross a wide street or open square? yes no
66. Does it make you uneasy to sit in a small room with the door shut? yes no
67. Do you usually know just what you want to do? yes no
68. Do you worry too much about little things? yes no

- | | | |
|--|-----|----|
| 69. Do you think you worry too much when you have an unfinished job on your hands? | yes | no |
| 70. Do you think you have too much trouble in making up your mind? | yes | no |
| 71. Can you do good work while people are looking on? | yes | no |
| 72. Do you get rattled easily? | yes | no |
| 73. Can you sit still without fidgeting? | yes | no |
| 74. Does your mind wander badly so that you lose track of what you are doing? | yes | no |
| 75. Does some particular useless thought keep coming into your mind to bother you? | yes | no |
| 76. Can you do the little chores of the day without worrying over them? | yes | no |
| 77. Do you feel you must do a thing over several times before you can drop it? | yes | no |
| 78. Are you afraid of responsibility? | yes | no |
| 79. Do you feel like jumping off when you are on high places? | yes | no |
| 80. At night are you troubled with the idea that somebody is following you? | yes | no |
| 81. Do you find it difficult to pass urine in the presence of others? | yes | no |
| 82. Do you have a great fear of fire? | yes | no |
| 83. Do you ever feel a strong desire to go and set fire to something? | yes | no |
| 84. Do you ever feel a strong desire to steal things? | yes | no |
| 85. Did you ever have the habit of biting your fingernails? | yes | no |
| 86. Did you ever have the habit of stuttering? | yes | no |
| 87. Did you ever have the habit of twitching your face, neck, or shoulders? | yes | no |
| 88. Did you ever have the habit of wetting the bed? | yes | no |
| 89. Are you troubled with shyness? | yes | no |
| 90. Have you a good appetite? | yes | no |
| 91. Is it easy to make you laugh? | yes | no |
| 92. Is it easy to get you angry? | yes | no |
| 93. Is it easy to get you cross or grouchy? | yes | no |
| 94. Do you get tired of people quickly? | yes | no |
| 95. Do you get tired of amusements quickly? | yes | no |
| 96. Do you get tired of work quickly? | yes | no |
| 97. Do your interests change frequently? | yes | no |
| 98. Do your feelings keep changing from happy to sad and from sad to happy without any reason? | yes | no |
| 99. Do you feel sad or low-spirited most of the time? | yes | no |
| 100. Did you ever have a strong desire to commit suicide? | yes | no |
| 101. Did you ever have heart-disease? | yes | no |

178 Diagnosing Personality and Conduct

102. Did you ever have St. Vitus's dance?	yes	no
103. Did you ever have convulsions?	yes	no
104. Did you ever have anemia badly?	yes	no
105. Did you ever have dyspepsia?	yes	no
106. Did you ever have asthma or hay fever?	yes	no
107. Did you ever have a nervous breakdown?	yes	no
108. Have you ever been afraid of going insane?	yes	no
109. Has any of your family been insane, epileptic, or feeble-minded?	yes	no
110. Has any of your family committed suicide?	yes	no
111. Has any of your family had a drug habit?	yes	no
112. Has any of your family been a drunkard?	yes	no
113. Can you stand pain quietly?	yes	no
114. Can you stand the sight of blood?	yes	no
115. Can you stand disgusting smells?	yes	no
116. Do you like outdoor life?	yes	no

A rough classification of these questions may be made as follows:

TABLE 24

CLASSIFICATION OF ITEMS IN THE WOODWORTH PSYCHONEUROTIC INVENTORY

Physical symptoms, pains, weariness, incoördinations....	28
Adjustment with the environment	20
Fears, worries	16
Unhappiness, unsocial and antisocial moods and conduct..	16
Dreams, phantasies, sleep disturbance	10
Reactions to drink, tobacco, drugs, sex	7
Mental symptoms	6
Vacillations	5
Compulsions	4
Questions about one's family	4

116

Just as intelligence tests were found to be useful in classifying and grouping school children on the basis of ability, so, it was suggested, perhaps the Woodworth Psychoneurotic Inventory might be serviceable in identifying children who have difficulty in making school or home adjustments. Accordingly in 1921 Miss Ellen Mathews (20) used the questionnaire with the purpose in mind of adapting it for school-children. The original Woodworth list was found unsuitable, however. Some of the questions were unintelligible to children. Others caused too great embarrassment, restlessness, or tittering in the group, and in one private school for girls the use of the questionnaire caused considerable excitement

among the parents. Mathews revised the questionnaire by dropping some questions and adding others, so that her list of seventy-five questions is one that may be used with children as young as twelve or thirteen years of age.*

Cady (5), compiling, in 1923, a group of tests to estimate juvenile incorrigibility, selected the Woodworth questionnaire as one of his battery. He argued that probably much juvenile delinquency was due to imperfect adjustment to the demands of life and that the Woodworth questions should aid in revealing the tendency to make faulty or inadequate adjustments. Cady selected questions from the Woodworth questionnaire and from another revision by Johnson (16) and added twelve questions which bear distinctly on the habits and expressions of the behavior of incorrigibles. Cady prepared an alternative form of his questionnaire by revising all questions so that an answer of *yes* in the first form would correspond to an answer of *no* in the second form and vice versa. This enabled him to determine the consistency of response, and also the reliability of the questionnaire; furthermore, by using the combined result of both forms the reliability was increased.

In Cady's report he states that his revised questionnaire consists of fifty-nine questions. Terman (29) prints a list purporting to be the questionnaire in his *Genetic Studies of Genius*, Volume I, containing eighty-five questions. Twelve of these are inserted for padding to lull the suspicions of the subject as to the purpose of the test. This revision, known as the Woodworth-Cady questionnaire, is the best one to use for children in their early teens.

The need for an instrument which will diagnose tendencies to make inadequate adjustments is especially needed on the college level. It is well known that many students, incapable of making satisfactory adjustments, are in need of guidance in mental hygiene. Laird (17), at Colgate University, appreciating this problem, tried out the Woodworth questionnaire on the college level and eventually published his own "Personal Inventory," a set of questions differing in some important particulars from other variations of the Woodworth Psychoneurotic Inventory. In the first place, Laird realized that those answering the questionnaire were not simply answering questions, but were engaged in a

*Published by C. H. Stoelting Company.

kind of self-rating. Since such judgments of oneself are capable of more than a dichotomous division into *yes* and *no* answers, he arranged that each item of his inventory should be responded to on a *graphic rating scale*. Second, he altered the scoring. On the basis of the replies of over 2,000 persons a distribution of the answers to each question was found. The quartile point toward the unfavorable end of each distribution was determined, and this quartile section was arbitrarily used as signifying that the reaction was unfavorable. Stencils were so prepared that a count could quickly be made of the number of responses that lay in the unfavorable area.

Third, Laird separated his questions into finer groups, called (I) Psychasthenoid (thirty-two items), (II) Schizoid (fourteen items), (III) Neurasthenoid (twenty-two items), and (IV) Hysteroid (seven items)—total, seventy-five items. Group I contains questions relating to obsessions, morbid fears, and doubts. An extreme form represented by those who answer a large number of the questions in group I unfavorably or who have a particularly acute form of one or more symptoms is known as *psychasthenia*. Group II contains questions relating mainly to social adjustment, and interest in persons and things. The implication is that persons who manifest poor adjustments of this type indulge excessively in day-dreams or phantasy thinking and hence have weakened their contacts with environmental realities. This represents an abnormal tendency which in the extreme leads to *schizophrenia* or *dementia praecox*. Group III contains questions that refer to physical conditions, especially aches, pains, and discomforts, fatigability, and anxiety about the health. These questions represent a tendency which, if extreme, is known as *neurasthenia*. Group IV is composed of questions relating to more pronounced physical disabilities such as fainting, paralysis, convulsions, loss of memory, etc. An extreme form of this tendency is known as *hysteria*. A revised form of the Colgate Personal Inventory omits the questions in Group IV.

House (15) also, on the basis of experiments with the Woodworth Psychoneurotic Inventory in colleges, presents a revision in which the items are chosen for power to differentiate between known normal and psychoneurotic subjects. In the final form, called the "Woodworth-House Mental Hygiene Inventory,"* he

* Published by C. H. Stoelting Company.

includes one innovation. It has long been a hypothesis of those who make case studies of mental abnormalities that such mal-functioning has its roots in childhood experiences. The psychoanalysts very definitely state that maladjustments in maturity may be traced to unfortunate childhood experiences. House, therefore, includes twenty-five questions relating to experiences which occurred before the age of fourteen, as well as fifty questions relating to experiences which occurred after the age of fourteen. In this questionnaire the general heading "This problem has occurred in my life" is followed by a list of items each of which is a brief statement of a personality adjustment problem. Each item is to be checked in one of three degrees: *extreme*, *moderate*, and *no*.

J. O. Chassell (6) in collaboration with Goodwin Watson, has constructed an "Experience Variables Record" which is an outgrowth of the Woodworth questionnaire. The innovations in this form are (a) an emphasis on the type of situation in which the maladjustment occurs rather than on the type of response itself; (b) an extension of the *rating* idea by describing degrees of the presence or absence of the variables in the questionnaire; and (c) an attempt to obtain data on childhood experiences as well as on present irritating situations. The record has twelve main sections:

- | | |
|--|--|
| 1. Mother Relationships | 8. Physical Development |
| 2. Father Relationships | 9. Intellectual Development |
| 3. Relationship with Brothers and Sisters | 10. Vocational Adjustment |
| 4. Family Life | 11. Social Situation—Adjustment in Comrade Groups, Status in Community, Public Recognition |
| 5. Religion and Standards | 12. General Emotional Adjustment, Happiness, etc. |
| 6. Sex Development | |
| 7. Love-Affairs—Crushes and Heterosexual Adjustments | |

These sections, Chassell believes, represent "the major situations in our society in which the growing youth finds it necessary to make adjustments and develop the more stable attitudes which constitute the framework of his character."

The questions in each section are divided into four groups as follows:

Group I. Environmental situations: general background.

Group II. Environmental factors bearing more directly upon the individual subject.

Group III. Subject's responses; his habits, interests, tendencies.

Group IV. Problems of adjustment and more "difficult situations" encountered by later adolescents.

Chassell and Watson, we repeat, place most of the emphasis on the situation, devoting less attention to the peculiar mode of response that the individual may have developed. In the blank they provide space after nearly every scale for checking answers for childhood, early teens, and recent or present times.

As a sample of the graded descriptive scale to be used in answering each question, the following is taken from the Mother Relationship Section.

Her tendency to exhibit physical affection toward subject.

Very demonstrative, marked exhibition of affection

Steady general affection—consistent but not marked

Physical demonstrations on rare occasions—eve of departure, etc.

Very undemonstrative—subject doesn't ever remember sitting on her lap. Affection indicated by acts, solicitations, etc. Never kissed husband or children.

A total score is not computed, since the "Experience Variables Record" is intended primarily for the use to be made of the answers to specific items.

The present author (27), acting on the hypothesis that adjustment is specific according to the stress of the situation in which a person finds himself, brought together a series of 175 questions, divided into the seven following sections, to form an "Adjustment Questionnaire" for use with high school pupils:

TABLE 25
SECTIONS IN SYMONDS ADJUSTMENT QUESTIONNAIRE
(27, p. 322)

	<i>No. of items</i>
Adjustment in Relation to the Curriculum.....	24
Adjustment in Relation to Social Life of the School.....	23
Adjustment in Relation to the Administration.....	14
Adjustment in Relation to the Teachers.....	33
Adjustment in Relation to the Other Pupils.....	33
Adjustment in Relation to the Home and Family.....	36
Adjustment in Relation to Personal Affairs.....	12
Total numbers of items	175

By tabulating the responses to each item made on the fifty highest-scoring papers out of a total of 162 and the fifty making the lowest score, the most significant items were found to be (27, p. 325):

Are you given a chance to tell or show what you know in your classes?

Do you think there should be more try-out or optional classes?

Are you required to take subjects that you dislike?

Would you select another teacher in any of your subjects if you were permitted to?

Do your teachers require too much homework?

Do your teachers make the assignments too long?

Do you like examinations in school?

Do your teachers usually understand your difficulties?

Do all of your teachers treat you as a friend?

Do any of your teachers have a wrong opinion about you?

Are you ever punished for things you do not do?

Do any of your teachers mark examinations too severely?

Do you feel that you are making quite a success of the things that you do?

Do you often fail in the subjects that you dislike?

Do all of your teachers make the assignment clear?

Do your teachers praise you when you hand in good work?

Do you think that any of your teachers are too strict?

Are you doing as much or as well in school as your parents expect you to do?

Do you think your work this year is rather monotonous?

Practically all of these items have to do with a pupil's success in school and out, since one may assume that being successful and being appreciated by others really do constitute the most important factors in determining the quality of adolescent adjustments.

For the time being, a high point in the development of adjustment questionnaires has been reached by the "Personality Schedule" developed by L. L. and T. G. Thurstone (30). This extensive inventory of 223 questions is drawn from the work of Woodworth, House, Laird, Freyd, and Allport. From the study of 694 University of Chicago freshmen the following items were determined to be the most differentiating:

Do you get stage fright?

Do you have difficulty in starting a conversation with a stranger?

184 Diagnosing Personality and Conduct

Do you worry too long over humiliating experiences?

Do you often feel lonesome, even when you are with other people?

Do you consider yourself a rather nervous person?

Are your feelings easily hurt?

Do you keep in the background on social occasions?

Do ideas often run through your head so that you cannot sleep?

Are you frequently burdened by a sense of remorse?

Do you worry over possible misfortunes?

Do your feelings alternate between happiness and sadness without apparent reason?

Are you troubled with shyness?

Do you day-dream frequently?

Have you ever had spells of dizziness?

Do you get discouraged easily?

Do your interests change quickly?

Are you easily moved to tears?

Does it bother you to have people watch you at work even when you do it well?

Can you stand criticism without feeling hurt?

Do you have difficulty in making friends?

Are you troubled with the idea that people are watching you on the street?

Does your mind often wander badly so that you lose track of what you are doing?

Have you ever been depressed because of low marks in school?

Are you touchy on various subjects?

Are you often in a state of excitement?

Do you frequently feel grouchy?

Do you feel self-conscious when you recite in class?

Do you often feel just miserable?

Does some particular useless thought keep coming into your mind to bother you?

Do you hesitate to volunteer in a class recitation?

Are you frequently in low spirits?

Do you often experience periods of loneliness?

Do you often feel self-conscious in the presence of superiors?

Do you lack self-confidence?

Do you find it difficult to speak in public?

Do you often feel self-conscious because of your personal appearance?

If you see an accident, are you quick to take an active part in giving help?

Do you feel you must do a thing over several times before you leave it?

Are you troubled with feelings of inferiority?

Do you often find that you cannot make up your mind until the time for action has passed?

Do you have ups and downs in mood without apparent cause?
Are you in general self-confident about your abilities?

Reliability. It is said that Woodworth obtained a reliability of .90 for his original questionnaire of 116 items. The present writer reports reliabilities of .90 and .84 in the Adjustment Questionnaire; Thurstone, a reliability of .946 for his Personality Schedule. Mathews reports .667 between the split halves; and .369 (boys) and .697 (girls) on retests. Cady reports a reliability of .55 on his original list and .49 and .47 on the revised list. House obtains correlations of .714 and .845 on retests. Hoitsma, in reporting on the Colgate Mental Hygiene Test, gives reliability coefficients of .85 for a repetition of the questionnaire and .79 for the split halves. Chassell reports the following coefficients in terms of Pearson C for the consistencies in answers between first and second responses to separate items in his questionnaire repeated after an interval of five to eight weeks.

TABLE 26
RELIABILITY COEFFICIENTS (PEARSON C) FOR CHASSELL EXPERIENCE
VARIABLES RECORDED

	75 men	75 women
Childhood674	.722
Early teens641	.719
Recent or present688	.700
Average668	.714

Chassell finds adults as consistent in answers regarding their childhood as regarding their present condition. One thing seems evident from these figures: the consistency of the questionnaire is greater for persons of maturity than for children. It seems reasonable to expect reliability coefficients of .50 to .60 from children twelve to fifteen years of age; of .70 to .80 from college students; and above .80 from still more mature individuals. This increase in reliability with increasing age may be explained as the growth of the ability of individuals to diagnose and judge their own states. Children are not used to noticing their own states and adjustments, and consequently fluctuate and give unreliable answers when asked about them. Because mature persons usually have given more thought to their adjustments, they are in a better position to report on them. Another reason for the lower re-

liabilities reported for younger subjects is the matter of reading difficulty. Where children do not understand the questions, their responses become more of a chance matter, and this tends to attenuate the reliability. The psychoneurotic inventory loses its usefulness below the high school level.

Validity. The Woodworth Psychoneurotic Inventory was originally intended for use in sorting men according to their ability to make satisfactory adjustments under stress. It might be called a measure of *adaptability*, meaning by adaptability the capacity to adapt oneself in satisfactory ways to unfamiliar and trying circumstances. That the questionnaire does this is attested to by several bits of evidence. Woodworth himself says that in a group of psychoneurotic patients the average score (unfavorable responses) was 36, while the average score for normal individuals was 10. Mathews (20) reports correlations of .515 and .663 between scores on the test and judgments by teachers of nervous stability. House (15) finds that there is a distinct difference in scores made by normal college students and psychoneurotic patients.

On the hypothesis that many of the behavior problems of children arise from poor adjustments of the type indicated by the psychoneurotic inventory, several investigators have used the questionnaire as a measure of juvenile delinquency, obtaining results which support the surmise. Cady (5) reports correlations of .41, .42, and .36 with teachers' estimates of incorrigibility, which, in view of his rather low reliability coefficients, indicates a pronounced relationship. Cushing and Ruch (7) report a biserial r of .60 on the questionnaire between a group of fifty delinquent girls and fifty normal girls. Slawson (25) finds that 84.4 per cent of delinquent boys make *higher scores* than the median of normal boys as measured by Mathews. Bridges and Bridges report that delinquent boys make an average of twenty-one unfavorable responses on the Mathews revision of the questionnaire against an average of nine found by Mathews for normal subjects.

Miscellaneous findings on the questionnaire help us to interpret its significance. Mathews reports a decrease in score with age, one reasonable explanation of which is that the unreliability of the questionnaire for children has a tendency to throw the unfavorable responses toward 50 per cent, which automatically raises the score. Hollingworth (14) found a large drop in the score on the

questionnaire immediately after the armistice that ended the World War, indicating that the tendencies which the questionnaire signifies are functional in nature and tend to be affected by the *present* irritating condition rather than by some *past* circumstance. Both Everett (9) and Mathews (20) report higher scores among girls than among boys.

This questionnaire has revealed distinct racial differences. One study (1) credits Poles with the highest scores, followed in order by the French, Czechs, and Americans. Another study showed negroes making the highest scores, with Italians, Hebrews, and Americans following. The interpretation here need not be that races have inherited tendencies toward maladjustment, but that the questionnaire reflects differences in the environment, general conditions of living, racial traditions, etc.

Landis, Gullette, and Jacobsen (19) attempted to determine the degree to which the Psychoneurotic Inventory measures emotionality. The correlation of the questionnaire with ratings of emotionality was .31, with ratings of emotional stability .21, and with ratings of expressiveness —.11. In the same study the correlation of the questionnaire with speed of tapping was .57 and with vocabulary test .44. The questionnaire has apparently only slight if any reference to emotionality.

Garrett and Kellogg (12) (reversing earlier findings of Naccarati and Garrett (22)) find no relation between the morphologic index and the Psychoneurotic Inventory.

Mathews (20) found that the questionnaire correlates —.201 with the IQ for boys and —.055, —.425, and —.592 for girls. Hoitsma (13) reports a correlation of the Colgate Mental Hygiene Questionnaire with the Thorndike Intelligence Test for High School Graduates of .008; also with scholarship of .074. Thurstone (30) finds a correlation of +.037 between the Personality Schedule and the American Council Intelligence Test for University of Chicago freshmen.

Bridges (2), in a stimulating study using the Woodworth questionnaire on college students, concludes that college students show poorer adjustment than the general population. The women students are less well adjusted than the men students. The typical student psychoneurosis is an anxiety neurosis; symptoms of hysteria, psychasthenia, and the major psychoses are rare. The most frequent symptoms are disturbed sleep, worry, irritability, per-

severation of ideas, and self-consciousness. There is no correlation with intelligence.

Both House (15) and Chassell (6) agree in finding a marked relationship between the items answered for the present and those answered for childhood. House makes his correlations on the basis of total score on his inventory and finds correlations of .705 and .803 between reported childhood reactions and present reactions. Chassell treats his items singly, using Yule's Q , and finds significant relations between items. These relationships should not be taken too seriously. However much we believe that present maladjustments are due to habits carrying over from childhood experiences, the relationships disclosed by these investigators may be explained more easily. Subjects answering the questions may actually believe that their present troubles began as a consequence of childhood experiences and so make their memories testify to their belief. Or, their memories being weak, they may not be able to make very sharp distinctions between present maladjustments and previous conditions. Answers to questions about childhood experiences have dubious value in any case.

In summary it can be said that the Woodworth Psychoneurotic Inventory and its later revisions have been found to indicate roughly the degree to which a person is making poor adjustments with irritating and difficult conditions of living. Various writers emphasize the fact that the inventory shows tendencies only and that persons with high scores should be subjected to a more thorough clinical examination. Others believe that a scrutiny of the individual items on the questionnaire may be used to make a diagnosis of the individual's peculiar form of maladjustment. The Woodward questionnaire has also been found to indicate behavior difficulties and has been suggested as a test of delinquent tendencies.

Further work with this type of questionnaire will undoubtedly follow along more special lines. Now that the general usefulness of this type of questionnaire has been demonstrated, effort should be made to use similar sets of questions to diagnose particular kinds of maladjustment. In the past psychopathology has limited its classifications to types of responses. But the more significant grouping may be around the types of situations that cause maladjustments. Different individuals find various methods of adjusting themselves to the same annoying situations.

Pressey X-O Tests

In 1919 Pressey (46) announced the construction of a new type of questionnaire which has subsequently had rather extended use. In Pressey's original statement concerning his test he seems more enthusiastic about the form of the test than about its content. At that time he believed that test-makers were constructing their tests so that they were needlessly artificial. He hit upon the scheme of inserting in a test list irrelevant elements which could be crossed out by the person taking the test. This could be applied to all forms of testing—intelligence, learning, reading, ingenuity, etc., as well as to tests for obtaining responses relative to beliefs, interests, and likes and dislikes. The feature of the test—*crossing out* irrelevant terms—has been symbolized in the name *X-O tests*.

In the original "Group Scale for Investigating the Emotions," (45) five different "tests" were included. In a revision, completed in 1920, there were changes in the lists and four different "tests" were included: One is a "test" in which the subject is instructed to cross out words whose meaning is unpleasant to him. Four types of unpleasantness are recognized in this test and words are included to diagnose abnormal fears, disgusts, abnormal sex tendencies, and abnormal self-regard.

The second "test" consists of words to be crossed out which are associated with a key word at the beginning of the line. This is a variation of the free association test.

In a third "test," words are to be crossed out which represent things that the subject considers wrong, making it a kind of ethical discrimination test.

In a fourth "test" words are to be crossed out which refer to things about which the subject has ever worried, the material being taken largely from Woodworth's Psychoneurotic Inventory. The words in each "test" have been carefully selected on the basis of experience with psychopathic patients.

Some difficulty was experienced in using these X-O tests in the public schools because of the unwillingness of teachers to use a "test" including words relative to sex and sexual conduct. Accordingly Test 1 containing words whose meaning is unpleasant, and Test 2 in which words associated with certain key words are to be crossed out were omitted; and a new "test" known as Form B

190 Diagnosing Personality and Conduct

was substituted in which the subject is instructed to cross out everything he likes or is interested in. All "tests" in this form consist of twenty-five lines of material.

The three "tests" begin as follows (34, p. 304):

Test 1

Directions.—Read through the twenty-five lists of words given just below and cross out *everything that you think is wrong*—everything that you think a person is to be blamed for. You may cross out as many or as few words as you like; in some lists you may not wish to cross out any words. Just be sure that you cross out everything you think is wrong.

1. begging, smoking, flirting, spitting, giggling;
2. fear, anger, suspicion, laziness, contempt;
3. dullness, weakness, ignorance, meakness, stinginess;
4. fussiness, recklessness, silliness, nagging, fibbing;
5. extravagance, sportiness, boasting, deformity, talking-back.

Test 2.

Directions.—Read through the twenty-five lists below and cross out *everything about which you have ever worried, or felt nervous or anxious*. You may cross out as many or as few words as you like; there may be some lines in which you may not wish to cross out any. But be sure you cross out everything about which you have ever worried.

1. loneliness, work, forgetfulness, school, blues;
2. sin, headache, fault-finding, sneer, depression;
3. meanness, clothes, sickness, looks, unfairness;
4. discouragement, self-consciousness, failure, accidents, worry;
5. temper, disease, pain, money, awkwardness.

Test 3.

Directions.—Read through the twenty-five lists just below and cross out *everything you like or are interested in*. You may cross out as many or as few words as you wish; there may be some lines in which you will not wish to cross out anything. But be sure you cross out everything that you like.

1. fortune-telling, boating, beaches, mountains, vaudeville;
2. camping, tennis, hiking, eating, amusement parks;
3. Beethoven, Edison, Napoleon, Raphael, Tennyson;
4. kissing, flirting, pretty girls, talkative girls, athletic girls;
5. studying, dancing, day-dreaming, walking, reading.

Three hundred and seventy-five *items* are included in this form, which covers only three pages. Twenty minutes' working time is allowed in the case of college students.

In the original "scale," Pressey used two methods of scoring, one of which he calls the "affectivity" or "emotionality" score; the other, the "idiosyncrasy" score. The "affectivity" score was obtained by merely totaling the number of words crossed out. In determining the "idiosyncrasy" score, the modal word crossed out in each line was first determined (in a group of 114 college students). The "idiosyncrasy" score is the number of words crossed out which are not the modal word for each line.

Pressey definitely presents these X-O tests as research instruments and not as finished products measuring anything in particular. In his original statement he spoke of an intention to measure "emotional instability," but this original aim was not kept foremost in subsequent revisions. Pressey says (45, p. 61), "The scores on the entire examination are the blurred result of a number of factors, and are of relatively little importance. However, it is possible, from the mass of data yielded by the examination, to combine certain items in such a way as to obtain, from the single examination, highly differential information with reference to a number of problems."

McGeoch and Whitely (41) have made an elaborate study of the reliability of the Pressey tests, using as subjects sophomores in Washington University. Three groups were given the tests with repetitions after forty-eight hours, forty-five days, and ninety days, respectively. Table 27 on the following page taken from their article tells the story.

The reliabilities for the forty-eight-hour interval are fairly satisfactory, but reliabilities tend to fall off for the longer periods. This is explained by the investigators as due to an actual change in "emotional organization."

The general impression which one receives from studying the reports of the use of the Pressey X-O tests is that the tests as a whole are a composite of several different kinds of tasks and that as a result the total affectivity score or the idiosyncrasy score does not measure any one thing with much success. Pressey frankly states in his announcements that he has assembled a miscellaneous group of material for experimental study. Another impression one receives is that such a high premium is placed on

TABLE 27

PRESSEY X-O TESTS—RELIABILITY COEFFICIENTS

(from McGeech and Whitely, 41, pp. 262-264)

AFFECTIVITY SCORES					
<i>N.</i>	<i>Interval</i>	<i>Test 1</i>	<i>Test 2</i>	<i>Test 3</i>	<i>Test 4</i>
64	48 hours	.85	.86	.82	.87
48	45 days	.58	.74	.80	.75
70	90 days	.70	.67	.65	.51

IDIOSYNCRASY SCORES				
<i>Interval</i>	<i>Test 1</i>	<i>Test 2</i>	<i>Test 3</i>	<i>Test 4</i>
48 hours	.70	.55	.77	.43
45 days	.53	.28	.46	.52
90 days	.46	.43	.45	.53

CLASSIFICATION SCHEME FOR TEST 1				
<i>Interval</i>	<i>Disgust</i>	<i>Fear</i>	<i>Sex</i>	<i>Self-feeling</i>
48 hours	.88	.82	.90	.77
45 days	.39	.57	.55	.71
90 days	.76	.63	.60	.70

CLASSIFICATION SCHEME FOR TEST 4					
<i>Interval</i>	<i>Paranoia</i>	<i>Neurotic</i>	<i>Shut-in</i>	<i>Melancholia</i>	<i>Hypo-chondriacal</i>
48 hours	.80	.74	.87	.82	.84
45 days	.79	.76	.79	.74	.77
90 days	.57	.31	.59	.49	.59

understanding the vocabulary of the tests that responses are apt to be more or less chance for dull or even normal adults and for children. Bridges and Bridges (32), for instance, report that test 1 behaves much like an intelligence test. Tjaden (49) reports that the scores on the Pressey test correlate with IQ.

The fact that the "test" bears the title "Group Scale for Investigating the Emotions" has led many persons to assume that it is a measure of "emotionality" or something similar. Landis, Gullette, and Jacobsen (39, p. 225) in their study in emotionality say, "Affectivity fails to give any significant correlations" and "It is hard to see why this idiosyncrasy rating should give other than chance correlations with other factors. The score here is based on the number of times the subject fails to agree with the dislike, association, blame, or worry which the modal person of a limited

standardizing group gave. The idea back of this test—namely, that the person who does not agree with the favorite worries or blames of his community is odd or peculiar—is correct. But in practice we believe it is impossible to standardize the test for all groups or communities. The worries, blames, and dislikes which an individual possesses are largely the product of his environment. The city man has one set of beliefs and standards, and the farmer another, and so on from one group to another. . . . We are of the opinion that the Pressey test embodies several very worth-while ideas, but should have a thorough revision and new standardization.”

Bridges and Bridges (32) and Lentz (40) have studied the reactions of delinquents to the Pressey X-O tests. All three agree that the Pressey tests do not differentiate, so far as score goes, between delinquents and non-delinquents. But still it was evident that delinquents and non-delinquents made different responses on separate items. Bridges and Bridges, for instance, found that delinquents considered fewer things wrong than college students, and likewise that they had more worries and fewer interests. The Woodworth-Mathews questionnaire is positively correlated with Pressey worries and interests, and negatively with Pressey things considered wrong.

Weber and Guilford (50) draw similar conclusions with regard to the use of the Pressey tests with criminals: that while they show marked divergence from the norm on separate elements, the total scores are not diagnostic of criminal tendencies.

Naccarati and Garrett (43) find that the Pressey test has no relationship with the morphologic index.

These negative results should be contrasted with the positive findings of Chambers in using the Pressey tests to measure “emotional maturity” (33) and “college achievement” (34). Chambers’ plan in both of these investigations was to study the differences in responses of groups known to differ markedly in a particular quality being studied. In the case of emotional maturity, for instance, he studied the responses of a group of 166 pupils in the sixth and eighth grades, and another group of 196 pupils in the tenth and twelfth grades. The responses of each group were recorded in terms of percentage of items marked, and all items in which the groups showed a difference of 15 per cent or over were taken to be significant of emotional maturity. Ninety-four such

194 Diagnosing Personality and Conduct

words constituted a differential unit, and the papers of these groups as well as additional groups from the fourth grade, the sixth grade, and college were rescored on the basis of these ninety-four items. These groups showed decreasing scores with increasing maturity, a finding which leads Chambers to claim that he has a measure of "emotional" development. He suggests that a child showing a score of twenty points' deviation on either side of the median for his group may be suspected of emotional maladjustment. However, one would have to know with what else this differential unit correlates before such a sweeping statement could be accepted. It may be suspected that part of the decrease in things considered wrong and things worried about among more mature pupils may be due to the better discernment which comes with intellectual development.

Chambers, using a similar technique, studied the words which showed a differentiation in response among college students with records of good and poor achievement. The significant words are given below (34, p. 307).

TEST I (THINGS CONSIDERED WRONG)

<i>Words marked more often by good students</i>	<i>Words marked more often by poor students</i>
---	---

sportiness
war
king
toughness
betting
shame
nerve
bribery
craps
overeating

fear
socialism
day-dreaming
slowness
spending
bashfulness
absent-mindedness
pedler
union

TEST II (THINGS WORRIED ABOUT)

<i>Words marked more often by good students</i>	<i>Words marked more often by poor students</i>
---	---

books
self-consciousness
accidents
rivals
parties

work
failure
police
wrecks
dreams

TEST III (THINGS LIKED)

<i>Words marked more often by good students</i>	<i>Words marked more often by poor students</i>
mountains	dancing
leaders	smoking
camping	rough boys
Napoleon	rich boys
day-dreaming	society
reading	bargains
teaching	
Mowgli	
D'Artagnan	
teachers	
books	

Chambers reports correlations as follows for fifty-seven cases (34, pp. 308, 309):

TABLE 28
CORRELATIONS BETWEEN PRESSEY X-O TEST AND INTELLIGENCE AND
SCHOOL MARKS
(from Chambers)

r grades—X-O46
r grades—intelligence53
r X-O—intelligence51
r grades—X-O, Part I.....	.16
r grades—X-O, Part II.....	.00
r grades—X-O, Part III.....	.42

We may conclude from the array of evidence that the Pressey X-O test contains material which may be of service in diagnosing various conduct trends. But exactly which elements these are can only be discovered by the type of analysis used by Chambers. Total scores on the Pressey X-O tests are such a complex of various sorts of reactions as to be indicative of nothing that can be simply and usefully designated.

Questionnaires to Measure Introversion-Extroversion

The questionnaire has also been found to be a convenient instrument for measuring the differences between individuals in respect to introversion-extroversion. In our discussion of the Woodworth Psychoneurotic Inventory, it was noted that whereas it was easy to describe certain outstanding *situations* that are baffling and vexing and that cause maladjustment, it was less

satisfactory to test the tendency to poor adjustments by asking subjects about the *responses* themselves, because different individuals have such varied methods of making these emergency adjustments. Now it can be added, however, that these responses can all be examined with respect to one particular feature of adjustment. Some individuals, who tend to turn themselves outward in their attempts at adjustment and to make vigorous even though ineffectual responses, are characterized by an abundance of energy and zeal and restlessness quite out of proportion to the results that their efforts bring. Other individuals, turning inward and tending to withdraw their contacts from the outside world, attempt to find their satisfactions in their own verbal and implicit reactions. The tendency to make one or the other of these types of adjustments may be discovered by certain well-placed questions. The distinction between these two types of adjustment in annoying situations was first elaborated by Jung in his *Psychological Types*.

Laird (67), Marston (70), and Heidbreder (63) have severally but so nearly simultaneously published accounts of their efforts to measure introversion and extroversion that it is difficult to assign the credit of being first to any one of them. Perhaps all these attempts are really derivatives of the Woodworth Inventory, because certain of the questions asked in the introversion-extroversion questionnaires were also used by Woodworth in his inventory. Woodworth, for instance, asked, "Do you make friends easily?" and Laird uses a similar question.

Laird (67) in the original Colgate Personal Inventory, had a section C-1 of fifty-three questions for measuring introversion-extroversion. The technique of the graphic rating scale, already described on page 158 for section B-1 of the Colgate Personal Inventory, was used. Laird, finding that the introversion-extroversion division of the Personal Inventory was the most valuable section, subsequently subjected it to thorough revision. Out of over 100 questions alleged to be indicative of introversion-extroversion, Laird and his workers selected forty-one items as valid. These have been incorporated in revised forms of the Personal Inventory known as C-2, in which the student rates himself, and C-3, in which the student is rated by instructors or by friends.

A slight change has been made in the form of the scale in C-2 and C-3. Instead of using an unbroken line for the graphic

rating, the line has been broken into ten segments. The scoring key indicates how many segments are included in the area which may be definitely called introversion on the basis of a method similar to that employed in the first edition. The upper quartile was found from the distribution of responses for each item obtained from individuals who rated themselves on the "Inventory," and the critical segment for the scale of each question to be included in the scoring key was that which corresponded to the quarter of most unfavorable answers. Laird states that this dichotomous method of scoring correlates .88 with the score obtained by assigning credits according to the fifth of each scale in which the rating falls. He prefers this new method both because of its simplicity and because he is really interested in spotting individuals with high introversion tendencies. Scores might also be obtained by determining whether an individual's rating lies beyond a critical point on the extroversion end of each question scale. Laird states that such extroversion scores correlate $-.92$ with the introversion scores.

Heidbreder (63) took Freyd's collection of fifty-four specific characteristics of introversion and worked them up into a questionnaire. Each item was to be rated by a + if the characteristic was one which an individual possessed; by a - if the characteristic did not apply; and by a ? if the individual was neutral with respect to the characteristic. In her investigation, Miss Heidbreder not only had students answer the questions with regard to themselves, but had each student secure ratings on himself by two of his friends. Each item was studied with respect to its relation to the total, and it was found that there was a high consistency between items. In the list which follows, the first thirty-one items are diagnostic both according to self ratings and associates' ratings, while the last six, although not diagnostic statistically, show a direction *tendency* in harmony with the list as a whole. The items are in order of diagnostic value (63, pp. 129-131).

Characteristics of

Introversion

1. Limits his acquaintances to a select few.
2. Feels hurt readily; apparently sensitive about remarks or actions which have reference to himself.
3. Is suspicious of the motives of others.

198 Diagnosing Personality and Conduct

4. Worries over possible misfortunes.
5. Indulges in self-pity when things go wrong.
6. Gets rattled easily; loses his head in excitement or moments of stress.
7. Keeps in the background on social occasions; avoids leadership at social affairs and entertainments.
8. Is critical of others.
9. Prefers to work alone rather than with people; prefers to work at tasks that do not bring him into contact with people.
10. Has ups and downs in mood without apparent cause.
11. Is meticulous; is extremely neat about his dress and painstaking about his personal property.
12. Blushes frequently; is self-conscious.
13. Pays serious attention to rumors.
14. Expresses himself better in writing than in speech.
15. Resists discipline and orders.
16. Limits his acquaintances to members of his own sex.
17. Avoids all occasions for talking before crowds. Finds it difficult to express himself.
18. Is a radical; wants to change the world instead of adjusting himself to it.
19. Is outspoken; says what he considers the truth regardless of how others may take it.
20. Introspects; turns his attention inward toward himself.
21. Prefers participation in competitive intellectual amusements to athletic games.
22. Is strongly motivated by praise.
23. Day-dreams.
24. Is selfish.
25. Dislikes and avoids any process of selling or persuading any one to adopt a certain point of view (except in the religious field).
26. Is sentimental.
27. Prefers to read a thing rather than experience it.
28. Is extremely careful about the friends he makes; must know a person pretty thoroughly before he calls him a friend.
29. Shrinks from actions which demand initiative and nerve.
30. Prefers to work things out on his own hook; hesitates to accept or give aid.
31. Talks to himself.
32. Derives enjoyment from writing about himself.
33. Keeps a diary.
34. Shrinks when facing a crisis.
35. If he unburdens at all, he does so only to close personal friends and relatives.
36. Is reticent and retiring; does not talk spontaneously.
37. Is creative of new and sometimes eccentric ideas and things.
38. Works by fits and starts.

39. Is a poor loser; considerably upset and indisposed after the loss of a competitive game.
40. Depreciates his own abilities, but assumes an outward air of conceit.
41. Is absent-minded.
42. Hesitates in making decisions on ordinary questions in the course of the day.
43. Believes in "mind cures"; accepts an idealistic philosophy.
44. Has ups and downs in mood with apparent cause.
45. Rewrites his social letters before mailing them.
46. Is slow in movement.
47. Is governed by reason rather than impulse or emotion. Is a good rationalizer, i.e., can give good reasons for his actions.
48. Admires perfection of form in literature.
49. Makes mistakes in judging the character and ability of others.
50. Is thrifty and careful about making loans.
51. Is effeminate (if a man).
52. Is persistent in his beliefs and attitudes.
53. Takes up work which requires painstaking and delicate manipulation.
54. Is conscientious.

Marston (70) studied introversion-extroversion in young children two to six years of age. His questionnaire, of necessity answered by parents or teachers instead of by the children themselves, consists of twenty items, each made up of two statements describing the opposite poles of a characteristic which is thought to be significant of the introversion-extroversion tendency. Either of the two statements is to be marked with two plus signs if the description of the characteristic definitely describes a child, and with one plus sign if the child merely inclines in that direction. Marston classifies the items of his scale as follows:

	<i>No. of items</i>
Social or Self-Attitudes	10
Energy Qualities	7
Emotional Tendencies	3

Guthrie (61) mentions other methods of measuring introversion-extroversion. One is a test of *campus information* or *gossip*, by which the degree to which "students were in touch with their human environment or were recluses" might be determined. Naturally such a test has a rather marked correlation with in-

200 Diagnosing Personality and Conduct

telligence. A second method is the free-association test. A third method for measuring extroversion is the degree to which a student approximates the composite judgment of his group in ranking his instructors in order of effectiveness as teachers.

Conklin (55) has also devised an interest questionnaire designed to test introvert-extrovert differences. Forty activities are listed, such as to play baseball, to hear a lecture on classical music, to talk with friends about hunting, to visit an automobile show, to talk with friends about literature, to read essays on literary criticism, etc. Twenty of these have been statistically determined to be significant of introvert tendencies, and twenty of extrovert tendencies. Each of these is to be rated on a nine-point scale. The test is scored by obtaining the ratio of the sum of the reactions to the extrovert items to the sum of the reactions to the introvert items.

Marston (70) also studied some performance tests of introversion-extroversion.

One important fact which stands out from an inspection of the distribution of scores on an introversion-extroversion questionnaire is illustrated by the following representative distribution taken from Miss Heidbreder's study.

TABLE 29
FREQUENCY DISTRIBUTION OF SCORES ON INTROVERSION-EXTROVERSION
QUESTIONNAIRE

Positive scores indicate a tendency toward introversion
(from Heidbreder)

<i>Class interval</i>	<i>f</i>
+ 20 + 24	1
+ 15 + 19	7
+ 10 + 14	14
+ 5 + 9	31
0 + 4	62
- 1 - 5	70
- 6 - 10	108
- 11 - 15	85
- 16 - 20	106
- 21 - 25	66
- 26 - 30	35
- 31 - 35	9
- 36 - 40	5
- 41 - 45	1
	<hr/> 600

People do not group themselves into the two extreme types. The distribution of individuals in introversion-extroversion follows the normal law of frequency. Few people are extremely introvert; few are extremely extrovert; most persons occupy an intermediate position, and for such people the term *ambivert* has been coined.

A second important fact is that a randomly chosen group of persons tend to answer more questions as extrovert than as introvert. To be introvert is more unusual than to be extrovert, and could probably be classed also as more abnormal.

In the third place, individuals tend to rate themselves as being more introvert than their friends rate them.

Reliability. Hoitsma (64) reports a correlation of .674 for a repetition of the Colgate Personal Inventory C-1 and .45 for one half against the other half. The correlation of introversion scores vs. extroversion scores was $-.36$, $-.22$, and $-.45$. On the revised form C-3, Laird (68) reports a correlation of .85 between the rating of two associates on a third person and .90 for the ratings of a person on himself at different times. Conklin (55) reports a reliability of .72 for the Colgate questionnaire; and Guthrie (61) reports a similar correlation of .60. It seems probable that Laird loses more reliability than he believes by his dichotomous method of scoring. Heidbreder reports correlations of .55 between self ratings and associates' ratings and .40 between associates' ratings. Marston reports an average correlation of .89 for the two halves of his scale and an average correlation of .71 between the ratings of two judges. Conklin reports reliability coefficients of .95 and .92 for split halves of his questions, corrected by the Spearman-Brown formula.

The evidence reported by Heidbreder of the high relationship between the answers to separate questions and the total score deserves mention here.

The conclusion is that the instruments for measuring introversion-extroversion have *very* high internal consistency, but suffer from the unreliability of the ratings. Research to date indicates that something real is being measured by the questionnaires and that success in the use of the instruments is hampered only by the difficulty which subjects and associates have in answering the questions correctly and without bias.

202. Diagnosing Personality and Conduct

Validity. The main line of evidence that introversion-extroversion questionnaires are measuring what they claim to be is afforded by internal consistency of the items. Heidbreder (63), Conklin (55), and probably Laird have checked each item against the whole list and find a significant relationship.

Hoitsma (64) and Conklin (55) report a negligible correlation between their own instruments and intelligence on the college level. Guthrie also reports a correlation of .01 between the Colgate Personal Inventory C-3 and intelligence. Conklin claims a correlation of +.21 between the Colgate Personal Inventory and the "American Council Intelligence Test," which, as it is based on only fifty-one cases, may probably safely be dismissed as a chance deviation from zero. The introversion-extroversion questionnaires, then, show negligible correlation with intelligence.

Hoitsma reports a correlation of +.35 between the Colgate Personal Inventory and college scholarship. Guthrie (61) finds the same variables to correlate only .11 at the University of Washington. This is an issue of extreme importance which should be cleared up by further experimental work.

Several investigators find differences between occupational groups with respect to introversion-extroversion as measured by a questionnaire. Laird (68) reports that foremen and executives are more extrovert than introvert, while inspectors, accountants, and research engineers are more introvert than extrovert. Nurses are extrovert as a group, Laird reports, while Pechstein finds that teachers are introvert. In this connection Pechstein writes (71, p. 196):

"General tendencies were noted for the student teachers to be more introverted than sophomores, and for the teachers to be more introverted than the student teachers. Other comparisons within these major groups revealed that the older teachers tended to be more introverted than the younger teachers, the teachers with college degrees to be more introverted than the teachers without degrees, and the brighter teachers (as measured by group mental tests) to be more introverted than the duller teachers. . . . Lack of social adjustment and adaptability is especially undesirable in teachers, yet this study indicates a selective process whereby the more introverted women tend to get into teaching and to stay in it longer.

"All the subjects were unmarried. The hypothesis is suggested that introversion in a woman means lesser likelihood of marriage. The prospect of marriage keeps out of the profession many who

might enter, and the call of marriage draws from the profession the less introverted members."

Conklin reports the following facts for professional groups who answered his introversion-extroversion questionnaire:

TABLE 30
GROUP DIFFERENCES IN EXTROVERSION-INTROVERSION INTEREST RATIOS
(from Conklin, 55, p. 34)

<i>Student group (above freshman year)</i>	<i>No.</i>	<i>Av.</i>	<i>S.D.</i>	<i>S.D._M</i>
Journalism majors (female)	20	117.5	40.4	9.0
Journalism majors (male)	15	100.9	35.0	9.0
English lit. majors (female)	22	97.1	24.1	5.1
Physical educ. majors (female)	13	76.4	16.2	4.5
Business adm. majors (male)	40	66.9	17.3	2.7
Pre-medical majors (male)	32	77.2	21.8	3.8
Pre-law majors (male)	18	79.9	20.4	4.8
<hr/>				
<i>Non-student groups</i>	<i>No.</i>	<i>Av.</i>	<i>S.D.</i>	<i>S.D._M</i>
Life insurance salesmen	21	66.4	16.5	3.6
Bank employees (male)	112	78.1	28.2	2.7
Bank employees (female)	23	94.6	24.3	5.1

In spite of a more recent statement from Laird's laboratory that there is no relation between introversion and vocational choice, it seems to us plain that the introversion-extroversion characteristic has some significance as a factor in vocational group differences. Devices for measuring it deserve consideration as techniques of vocational guidance for mature persons, and meanwhile of course they must also be subjected to further investigation. If the results already reported are substantiated, we have in such questionnaires the promise of effective instruments for advising persons as to the type of work for which they are best fitted and in which they will be most contented. These questionnaires deserve also to find a place in educational guidance. As Conklin points out, taken with an intelligence test, they should assist pupils in choosing curricula wherein lie their greatest hopes for making successful adjustments.

Holtsma reports a correlation of .49 between form B and C-1 of the Colgate Mental Hygiene Test. Schwegler (73) reports a similar correlation of .50. This substantiates the claim of mental hygienists that psychoneurotic persons tend to make introvert adjustments.

Laird (68) claims a correlation of .60 between extroversion-introversion and the morphologic index, a relationship which, although higher than a more extensive investigation would reveal, indicates that possibly the tendency to make one or another type of adjustment is organic and is conditioned by the balance of glandular secretion.

Laird (68) makes the statement that women are more introvert than men, while Marston finds girls more introvert than boys. Heidbreder (62) could discover little difference between men and women in total score on her questionnaire, but reports that the two sexes make different responses on separate items. Certain items are more diagnostic of introversion for men than for women (62, p. 58).

"It will be observed, too, that most of the traits which are more characteristic of women—and which are less frequent among men—are those which would interfere with efficient work; while those which are characteristic of men—and which occur less frequently in women—are those which would keep an individual from being socially agreeable." . . . "It may not be amiss to observe that there is evidence in the psychological literature that character and personality traits are susceptible to training, and this evidence, together with the fact that the sexes are not expected to do the same kinds of work in the world, and that different modes of behavior are considered appropriate for men and women, makes it possible to explain the differences that appear without assuming any fundamental, native sex differences in temperament."

Miss Heidbreder concludes that "sex differences and introversion-extroversion differences act as independent variables."

Using the Marston Scale, Caldwell and Wellman (53) determined that extroversion characterizes certain types of leadership.

"Extroversion among the girls was most marked in the Science-Club chairman, student-council members, and magazine-staff members. In all types of leadership except athletics the girls were ranked as extroverts. The athletic leaders were ranked at balance between extroversion and introversion. The boys tended to be more extrovertive than introvertive, but not to such a marked degree as the girls. The two magazine-staff representatives were notable exceptions, ranking as decided introverts."

Schwegler (73) gave an extensive battery of tests to high school students chosen for being definitely introvert or extrovert on the showing of a questionnaire resembling that devised by Marston.

The biserial r expressing the degree to which these two contrasting groups differ is as follows:

TABLE 31

BISERIAL r BETWEEN INTROVERSION-EXTROVERSION DIFFERENCE AND VARIOUS FACTORS

A positive coefficient indicates a higher distribution of scores by introverts
(from Schwegler)

Age	-.031
McCall-Multimental T-score	-.339
Free-Association Test—mean time.....	.620
90 per cent time494
response failure198
reproduction failure	-.037
word range107
contrast responses	-.406
Ink blot—nouns	-.386
verbs111
Motor output—unspeeded	-.250
speeded	-.328
total	-.309
kinetic reserve054
Multiple choice257
Weight-selection054
suggestibility	-.216
Motivated choice time159
Picture suggestibility	-.061
Color-naming—first trial148
second trial108
total145
Trend (psychoneurotic) questionnaire496
Conditioned affect (Travis)	-.313

Schwegler characterizes an introvert as being

“in the presence of single familiar uncomplicated situations

- (a) slower in verbal response
- (b) less freely productive of words, ideas, and movements
- (c) slightly more tenacious in holding to the evidence of his own experience
- (d) less given to superficial automatized responses
- (e) more inclined to morbid anxieties, to autistic trends, and to psychasthenia, obsessions, and phobias, and
- (f) less inclined to admit the presence of a rich emotional life than is the contrasted extrovert.”

Ascendance-Submission

The tendency to dominate or submit to others is another characteristic which may yield to attack by the questionnaire

206 Diagnosing Personality and Conduct

self-rating method. *A priori* it appears that any individual has enough consistency in his behavior in various situations with regard to his dominance or shyness to make it worth while to attempt to measure it. G. W. Allport, together with F. H. Allport, studied the problem, assembled a series of situations in which persons exhibit ascendant or submissive behavior, and finally constructed questions by means of which a person could rate himself on this trait. A sample of the questions follow (78, p. 125):

"At church, a lecture, or an entertainment, if you arrive after the program has commenced and find that there are people standing but also that there are front seats available which might be secured without 'piggishness' or discourtesy but with considerable conspicuousness, do you take the seats?

habitually _____
occasionally _____
never _____"

Two forms of the questionnaire were constructed—one for men and one for women. Many of the questions bear close resemblance to questions found in introversion-extroversion questionnaires, indicating that this may be another expressive phase of some underlying constitutional difference.

The derivation of the scoring scheme is interesting and instructive. For all students who took the original test, ratings were also obtained (one by self and four by classmates) on ascendance-submission. These five ratings were then averaged for each individual to yield a criterion. An average of these criterion ratings was then obtained for each response made on the test.

TABLE 32
EVALUATION OF SINGLE ITEM ON ALLPORT ASCENDANCE-SUBMISSION QUESTIONNAIRE SHOWING DERIVATION OF SCORING SCHEME
(from Allport, 78, p. 128)

Choice	Average criterion	Average of all criterion scores	Difference
Habitually	3.35	3.48	+ .13
Occasionally	3.50	3.48	— .02
Never	3.57	3.48	— .09

The average of all criterion ratings was 3.48. The difference between those obtained for any response and the average of all taken to one decimal place, was used as the scoring value of the

item. No attempt was made to determine whether the differences in average criterion values for different responses were statistically significant, but since larger differences resulted in larger scoring values for a response, this matter was somewhat automatically cared for.

The reliability of the questionnaire was found to be .78 for women who repeated the questionnaire, and .58 (corrected to .74 by the Spearman-Brown formula) for men by the split-half method.

The questionnaire results were correlated against the ratings as a measure of validity. For the composite ratings this correlation was .586 and .496, for self-ratings alone .633, and for associates' ratings .459. But these correlations are spuriously high, since they were determined on the same groups that helped determine the scoring key. On fresh groups correlations of .29, .30, and .33 were obtained.

The authors of the test point out its possible usefulness in personal counseling, educational and vocational guidance, and vocational placement. The suggestion is made that possibly the questionnaire measures one factor in leadership.

The question as to whether or not the test artificially creates by its very name and score a distinction which has no real existence must be constantly borne in mind. It is quite possible that the tendency to dominate or submit is specific for each of the myriad social situations in which a person finds himself, and that the responses may actually bear little resemblance to one another. However, the very reliability of the questionnaire indicates an internal consistency of the items which is significant. It may be that this consistency simply reflects the consistency with which the person considers himself, as indicated by his replies to the questions, but even if this should be all its means, this fact would of itself be significant. However, the correlation with the ratings indicates that it corresponds to actual behavior characteristics which others can also observe.

The Wish

The wish as a diagnostic instrument possesses considerable significance, as is shown by preliminary studies by Washburne. By merely giving children an opportunity to state three wishes in various ways, Washburne (80) is able to make significant in-

208 Diagnosing Personality and Conduct

terpretations concerning adjustment. This practically untouched method is deserving of further trial and experimentation.

An integration of these various questionnaires designed to measure adjustment has been effected in a "Personality Inventory" constructed by Bernreuter.* This questionnaire of 125 questions to be answered by encircling one of the three responses *yes, no* and *?* is scored in four ways to yield (1) a measure of neurotic tendency, (2) a measure of self-sufficiency, (3) a measure of introversion-extroversion, (4) a measure of dominance-submission. Reliability coefficients are reported for college psychology classes all of which are over .85. The various scores have been validated by correlating them against (a) the Thurstone Neurotic Inventory, (b) the Bernreuter Self-Sufficiency Test, (c) the Laird C-2 Introversion Test, (d) the Allport Ascendance-Submission Reaction Study. The coefficients of correlation, when corrected for attenuation, are all over .90 (with one exception) and most of them very close to 1.00. An interesting set of intercorrelations indicates that the scale of neurotic tendencies and the scale of introversion-extroversion are closely related.

TABLE 33
INTERCORRELATIONS BETWEEN MEASURES OF NEUROTIC TENDENCY, SELF-SUFFICIENCY, INTROVERSION-EXTROVERSION, AND ASCENDANCE-SUBMISSION ON THE BERNREUTER PERSONALITY INVENTORY

(from Bernreuter)

	<i>Neurotic</i>	<i>Self-sufficiency</i>	<i>Introversion-extroversion</i>	<i>Ascendance submission</i>
Neurotic tendency		-.39	.93	-.82
Self-sufficiency			-.28	.50
Introversion-extroversion..				-.73

Summary

The questionnaire has been found useful as a measure of mal-adjustment. The psychoneurotic inventory, as devised by Woodworth, has proved its value as a measure of adaptability. It differentiates between normal and neurotic subjects, and between delinquents and non-delinquents. It has demonstrated its value in determining the degree of adjustment of school children with different phases of their environment. Questionnaires have also

* Published by the Stanford University Press, 1931.

been constructed which measure tendencies toward introversion-extroversion, ascendancy-submission, and other fundamental personality differences. The wish is another form of response which possesses considerable significance for adjustment. These questionnaires promise to be of service in vocational placement, in studying the adequacy of educational procedures, and in determining by survey methods those individuals whose maladjustment with their surroundings is serious enough to warrant further individual counsel and advice.

REFERENCES

The Psychoneurotic Inventory

1. ABEL, T. M., "National Differences and Preferences in Beliefs; A Statistical Study," *Journal of Comparative Psychology*, 6: 21-41 (1926).
2. BRIDGES, J. W., "Emotional Instability of College Students," *Journal of Abnormal and Social Psychology*, 22: 227-234 (1927).
3. ———, "A Study of a Group of Delinquent Girls," *Pedagogical Seminary and Journal of Genetic Psychology*, 34: 187-204 (1927).
4. BRIDGES, J. W., and BRIDGES, K. M. B., *A Psychological Study of Juvenile Delinquency by Group Methods*, "Genetic Psychology Monographs 1, No. 5 (Sept., 1926), 411-506.
5. CADY, V. M., *The Estimation of Juvenile Incurability*, *Journal of Delinquency Monographs*, No. 2 (1923), 140 pp.
6. CHASSELL, J. O., *The Experience Variables; a Study in the Variable Factors in Experience Contributing to the Formation of Personality* (published by the author, Rochester, N. Y., 1928).
7. CUSHING, H. M., and RUCH, G. M., "An Investigation of Character Traits in Delinquent Girls," *Journal of Applied Psychology*, 11: 1-7 (Feb., 1927).
8. ELWOOD, R. H., "The Rôle of Personality Traits in Selecting a Career; the Nurse and the College Girl," *Journal of Applied Psychology*, 11: 199-201 (June, 1927).
9. EVERETT, E. M., *Tests of Emotional Stability as Applied to High School Students*, unpublished master's essay, Department of Psychology, Columbia University, 1919.
10. FLEMMING, E. G., *The Predictive Value of Certain Tests of Emotional Stability as Applied to College Freshmen*, *Archives of Psychology*, No. 96 (1928).
11. FRANZ, S. I., *Handbook of Mental Examination Methods* (The Macmillan Company, New York, 1919), 193 pp.

12. GARRETT, H. E., and KELLOGG, W. N., "The Relation of Physical Constitution to General Intelligence; Social Intelligence and Emotional Instability," *Journal of Experimental Psychology*, 11: 113-129 (1928).
13. HOITSMA, R. K., "The Reliability Relationships of the Colgate Mental Hygiene Test," *Journal of Applied Psychology*, 9: 293-303 (1925).
14. HOLLINGWORTH, H. L., *The Psychology of Functional Neuroses* (D. Appleton and Company, New York, 1920), 259 pp.
15. HOUSE, S. D., *A Mental Hygiene Inventory*, Archives of Psychology, No. 88 (May, 1927).
16. JOHNSON, B., "Emotional Instability in Children," *Ungraded*, 5: 73-79 (1920).
17. LAIRD, D. A., "Detecting Abnormal Behavior," *Journal of Abnormal and Social Psychology*, 20: 128-141 (1925).
18. ———, "Mental Hygiene and Vocational Tests," *Journal of Educational Psychology*, 16: 419-422 (1925).
19. LANDIS, C., GULLETTE, R., and JACOBSEN, C., "Criteria of Emotionality," *Pedagogical Seminary and Journal of Genetic Psychology*, 32: 209-234 (1925).
20. MATHEWS, E., "A Study of Emotional Stability of Children," *Journal of Delinquency*, 8: 1-40 (1923).
21. McDOWELL, E. D., *Mental and Emotional Adjustments of Stuttering Children*, Teachers College Contributions to Education, No. 314 (1928).
22. NACCARATI, S., and GARRETT, H. E., "The Relation of Morphology to Temperament," *Journal of Abnormal and Social Psychology*, 19: 254-263 (1924).
23. SCHWEGLER, R. A., *A Study of Introvert-Extrovert Responses to Certain Test Situations*, Teachers College Contribution to Education, No. 361 (1929).
24. SLAWSON, J., *The Delinquent Boy—A Socio-Psychological Study* (Richard G. Badger, The Gorham Press, Boston, 1926).
25. ———, "Psychoneurotic Responses of Delinquent Boys," *Journal of Abnormal and Social Psychology*, 20: 261-281 (1925).
26. SUNNÉ, D., "Personality Tests: White and Negro Adolescents," *Journal of Applied Psychology*, 9: 256-280 (1925).
27. SYMONDS, P. M., "An Adjustment Survey," *Journal of Educational Research*, 21: 321-330 (May, 1930).
28. TENDLER, A. D., "Measurable Factors of Adjustment in Juvenile Delinquents," *Jewish Social Service Quarterly*, 1: 1-31.
29. TERMAN, L. M., *Genetic Studies of Genius*, Vol. I (Stanford University Press, 1925), 648 pp.
30. THURSTONE, L. L., and THURSTONE, T. G., "A Neurotic Inventory," *Journal of Social Psychology*, 1: 3-30 (Feb., 1930).

Pressey X-O Test

31. BOND, H. M., "Non-Intellectual Traits of Negro Adults," *Journal of Abnormal and Social Psychology*, 21: 267-276 (Oct., 1926).
32. BRIDGES, J. W., and BRIDGES, K. M. B., *A Psychological Study of Juvenile Delinquency by Group Methods*, Genetic Psychology Monographs, Vol. I, No. 5 (Sept., 1926), 411-506.
33. CHAMBERS, O. R., "A Method of Measuring the Emotional Maturity of Children," *Pedagogical Seminary*, 32: 637-647 (1925).
34. ———, "Character Trait Tests and the Prognosis of College Achievement," *Journal of Abnormal and Social Psychology*, 20: 303-311 (1925).
35. ———, "Measurement of Personality Traits," in PRESSEY, S. L., PRESSEY, L. C., and others, *Research Adventures in University Teaching: Eighteen Investigations Regarding College and University Problems* (Public School Publishing Company, 1927), 152 pp.
36. COLLINS, M., "Character and Temperament Tests," *British Journal of Psychology*, 16: 89-99 (1925).
37. FLEMMING, E. G., *The Predictive Value of Certain Tests of Emotional Stability as Applied to College Freshmen*, Archives of Psychology, No. 96 (1928).
38. GUILFORD, J. P., "An Attempted Study of Emotional Tendencies in Criminals," *Journal of Abnormal and Social Psychology*, 21: 240-244 (Oct.-Dec., 1926).
39. LANDIS, C., GULLETTE, R., and JACOBSEN, C., "Criteria of Emotionality," *Pedagogical Seminary and Journal of Genetic Psychology*, 32: 209-234 (1925).
40. LENTZ, T. F., JR., *An Experimental Method for the Discovery and Development of Tests of Character*, Teachers College Contributions to Education, No. 180 (1925), 47 pp.
41. MCGEOCH, J. A., and WHITELEY, P. C., "The Reliability of the Pressey X-O Test for Investigating the Emotions," *Pedagogical Seminary and Journal of Genetic Psychology*, 34: 255-270 (1927).
42. MULLER, H. J., "Mental Traits and Heredity," *Journal of Heredity*, 16: 433-448 (1925).
43. NACCARATI, S., and GARRETT, H. E., "The Relation of Morphology to Temperament," *Journal of Abnormal and Social Psychology*, 19: 254-263 (1924).
44. OLSON, E. M., "A Study of the Emotions in Psychopathic Personalities," unpublished master's thesis, Department of Psychology, Columbia University, 1923.
45. PRESSEY, S. L., "A Group Scale for Investigating the Emo-

- tions," *Journal of Abnormal and Social Psychology*, 16: 55-64 (April, 1921).
46. PRESSEY, S. L., and PRESSEY, L. W., "'Cross-out' Tests, with Suggestions as to a Group Scale of the Emotions," *Journal of Applied Psychology*, 3: 138-150 (1919).
 47. PRESSEY, S. L., and CHAMBERS, O. R., "First Revision of a Group Scale Designed for Investigating the Emotions, with Tentative Norms," *Journal of Applied Psychology*, 4: 97-104 (1920).
 48. SUNNE, D., "Personality Tests: White and Negro Adolescents," *Journal of Applied Psychology*, 9: 256-280 (1925).
 49. TJADEN, J. C., "Emotional Reactions of Delinquent Boys of Superior Intelligence Compared with Those of College Students," *Journal of Abnormal and Social Psychology*, 21: 192-202 (July, 1926).
 50. WEBER, C. O., and GUILFORD, J. P., "Character Trends vs. Mental Deficiency in Delinquency," *Journal of Criminal Law and Criminology*, 16: 610-612 (1926).
 51. ZEIGLIN, L. H., "A Study of X, Psychometric and Otherwise," *American Journal of Psychiatry*, 1: 199-210 (1921).

Introversion-Extroversion

52. BROOM, M. E., "Critical Study of a Test of Extroversion-Introversion Traits," *Journal of Juvenile Research*, 13: 104-123 (April, 1929).
53. CALDWELL, O. W., and WELLMAN, B., "Characteristics of School Leaders," *Journal of Educational Research*, 14: 1-20 (June, 1926).
54. CONKLIN, E. S., "The Definition of Introversion, Extroversion and Allied Concepts," *Journal of Abnormal and Social Psychology*, 17: 367-382 (1923).
55. ———, "The Determination of Normal Extravert-Introvert Interest Differences," *Pedagogical Seminary and Journal of Genetic Psychology*, 34: 28-37 (1927).
56. DAVENPORT, M., "Inspectors and Foremen Compared on the Colgate Emotional Inventories," *Industrial Psychology*, 2: 263-264 (May, 1927).
57. ELWOOD, R. H., "The Rôle of Personality Traits in Selecting a Career; the Nurse and the College Girl," *Journal of Applied Psychology*, 11: 199-201 (June, 1927).
58. FLEMMING, E. G., *The Predictive Value of Certain Tests of Emotional Stability as Applied to College Freshmen*, Archives of Psychology, No. 96 (1928).
59. FLÜGEL, J. C., and RADCLIFFE, E. J. D., "Pressey Cross-Out Tests Compared with a Questionnaire," *British Journal of Medical Psychology*, 8: 112-131 (1928).

60. FREYD, M., "Introverts and Extroverts," *Psychological Review*, 31: 74-87 (1924).
61. GUTHRIE, E. R., "Measuring Introversion and Extroversion," *Journal of Abnormal and Social Psychology*, 22: 82-88 (1927).
62. HEIDBREDER, E., "Introversion and Extroversion in Men and Women," *Journal of Abnormal and Social Psychology*, 22: 52-61 (1927).
63. ———, "Measuring Introversion and Extroversion," *Journal of Abnormal and Social Psychology*, 21: 120-134 (July-Sept., 1926).
64. HOITSMA, R. K., "The Reliability and Relationships of the Colgate Mental Hygiene Test," *Journal of Applied Psychology*, 9: 293-303 (Sept., 1925).
65. JUNG, C. G., *Psychological Types*, translated by H. G. Baynes (Harcourt, Brace and Company, 1923), 654 pp.
66. LAIRD, D. A., "Mental Hygiene and Vocational Tests," *Journal of Educational Psychology*, 16: 419-422 (1925).
67. ———, "Detecting Abnormal Behavior," *Journal of Abnormal and Social Psychology*, 20: 128-141 (1925).
68. ———, "How Personalities Are Found in Industry," *Industrial Psychology*, 1: 654-662 (Oct., 1926).
69. ———, "Sex Indulgence and Psycho-neurotic Tendencies in Middle Adolescence," *Psychoanalytic Review*, 13: 496-498 (1926).
70. MARSTON, L. R., *The Emotions of Young Children*, University of Iowa Studies in Child Welfare, Vol. III (1925), 99 pp.
71. PECHSTEIN, L. A., "Introversion as a Factor in the Vocational Selection of Teachers," paper presented at the thirty-sixth annual meeting of the American Psychological Association, Columbus, Ohio, December, 1927; *Psychological Bulletin*, 25: 196, 197 (1928).
72. POWERS, N. E., "An Application of the Marston Introversion-Extroversion Rating Scale," *Journal of Educational Psychology*, 19: 168-174 (March, 1928).
73. SCHWEGLER, R. A., *A Study of Introvert-Extrovert Responses to Certain Test Situations*, Teachers College Contributions to Education, No. 361 (1929).
74. STEEN, F. H., and ESTABROOK, G. H., "The Relation between Introversion and Vocational Choice," *The Vocational Guidance Magazine*, 6: 327-328 (April, 1928).
75. STUART, J. C., "Data on the Alleged Psycho-pathology of the Only Child," *Journal of Abnormal and Social Psychology*, 20: 441 (Jan., 1926).
76. WEBER, C. O., and MAJGREN, R., "Experimental Differentia of Introversion and Extroversion," *Pedagogical Seminary*

214 Diagnosing Personality and Conduct

- and Journal of Genetic Psychology*, 36: 571-580 (Dec., 1929).
77. WELLS, F. L., "A Note on Numerical Score for Introvert and Extrovert Self-Rating," *Journal of Abnormal and Social Psychology*, 21: 316-318 (Oct.-Dec., 1926).

Ascendance-Submission

78. ALLPORT, G. W., "A Test for Ascendance-Submission," *Journal of Abnormal and Social Psychology*, 23: 118-136 (1928).
79. BENDER, I. E., "Ascendance-Submission in Relation to Certain Other Factors in Personality," *Journal of Abnormal and Social Psychology*, 23: 137-143 (1928).

The Wish

80. WASHBURN, J. N., "An Experiment in Character Measurement," *Journal of Juvenile Research*, 13: 1-18 (1929).

Chapter VI

ATTITUDE QUESTIONNAIRES

PROLIFIC use of the questionnaire has been made to explore and to tap attitudes on various social and ethical issues. Used in this way, the questionnaire takes the form of a rather lengthy ballot, and indeed it is so called in many of the studies. In describing the work done with this form, we shall make no attempt to present the studies in chronological order and no pretense to completeness. Several studies came out at about the same time, and much of the work has never been published. Our aim will be to describe those studies that offer the best suggestions for methodology and which yield critical data.

The present author constructed a sample ballot of 115 items on current social issues. These items were in the form of questions, each of which was to be answered by *yes* or *no*. Samples are (20, p. 316):

1. Is it desirable that schools be permitted which are conducted in a foreign language?
9. Should automobile drivers be given licenses without examination?
34. Should society deny any man the right to work?
59. Should the feeble-minded be educated?
85. Should the city maintain playgrounds?
100. Should an accurate record of births, marriages, and deaths be kept by a public agency?

This questionnaire was submitted to five persons: a sociologist, an English professor, two psychologists, and the writer, who were asked to answer each question according to what they believed to be the liberal, progressive, or radical position. *Liberal*, *progressive*, and *radical* were not defined except as being the opposite of *conservative*. It is not thought that these three terms are synonymous, but that they contain in common a point of view which would tend to produce a certain type of answer to each

question. Questions which the judges answered with three *yes*'s and two *no*'s or two *yes*'s and three *no*'s were thrown out, leaving 115 for which the issue was considered definite. After being worded so that there was an equal number of *yes* and *no* liberal answers, the 115 were then placed in random order. It is not contended that there is a *right* or *wrong* answer to these questions, but for a key the answers which were given as liberal, radical, or progressive were used. An impressionistic answer rather than a reasoned-out answer was desired from those taking the questionnaire. The score is the percentage of liberal answers out of questions tried.

Moore (12) used a similar twenty-item questionnaire on social issues, each item a question to be answered by *yes* or *no*. No statement is given as to how the "key" for this questionnaire was constructed except that it was based on the judgment of Moore, who studied the questionnaire, and of Rice, who constructed it.

Watson (30) in his "Survey of Public Opinion on Some Religious and Economic Issues," described in *The Measurement of Fair-mindedness*, has produced types of attitude measurement that will stimulate further productive work in this field for years to come. His is a set of tests which deserves careful study and analysis. As a total test the Watson survey may be used to determine the degree to which an individual tends to lean toward extreme opinions in either the radical or conservative direction. It measures the tendency to be extreme or biased or prejudiced in one's attitude or opinion toward social issues as opposed to being impartial or unprejudiced or "fair-minded"—hence the name "A Measure of Fair-mindedness." Besides this, the test makes possible a determination of the strength of prejudice in twelve different economic or religious directions, as shown in the following list:

1. Economic radicals.
2. Economic liberals.
3. Economic capitalists.
4. Persons fighting for a "social gospel," rather than an individual interpretation.
5. Persons interested mainly in a "personal gospel," prayer, mysticism, communion, salvation, etc.
6. Fundamentalists, orthodox "Apostles' Creed" variety.
7. Modernists, holding liberal Christian views.

8. Religious radicals, very broad, displeased with most existing Christian manifestations of religion.
9. Protestants who are inclined not to like Catholics.
10. Catholics who are inclined not to like Protestants.
11. Persons with high, strict standards of sex-ethics, or amusement, or "bad habits," or similar moral matters.
12. Persons with broad, loose standards of sex-ethics, or amusement, or "bad habits," or similar moral matters.

For purposes of description let us narrow our scrutiny to a point where we will consider the "Survey" merely as a measure of "economic radicalism," choosing all illustrations to represent this particular issue.

In Form A (test 1), "Cross-Out Test," a list of words is given. The directions are to cross out the words which suggest more that is disagreeable than that is agreeable. Examples of such words are:

Big Interests, Capitalist, Ku Klux Klan, Wall Street, Landlord.

The idea for this test is taken from the Pressey Cross-Out Tests. The words given as samples represent to the average man some of the most irritating elements in economic life. It is assumed that if a person crosses out any of these items he *tends* to oppose the established economic order. It presumably indicates a tendency to favor a change from the present state of economic affairs.

In Form B (test 2), "Degree of Truth Test," and in Form F (test 6), "Generalizations Test," statements are given which the subject is to approve or disapprove. In Form B the directions state that he is to indicate the *degree of truth* of the statement. If the response were merely to indicate whether the statement is true or false, this would become a true-false test, and as such it would not differ markedly in principle from the questionnaires by Symonds and Moore described above. In this particular instance, however, since Watson wishes his subjects to estimate degrees of truth or falsity, he puts before each question the symbols $+2$ $+1$ 0 -1 -2 one of which is to be encircled to designate that the statement is *utterly and unqualifiedly true*; *probably true*, or *true in large degree*; an *unchecked, open question*; *probably false*, or *false in large degree*; *utterly and unqualifiedly false*. Examples are:

218 Diagnosing Personality and Conduct

- +2 +1 0 -1 -2 The churches are more sympathetic with capital than with labor.
- +2 +1 0 -1 -2 To have experienced business men, who have made a financial success in private enterprise, hold the public offices of the country would be better than the present personnel.
- +2 +1 0 -1 -2 Unless industrial and economic conditions in the United States are remedied by sweeping changes in the present capitalistic system, we shall have a class revolution.

In Form F, "Generalization Test," similar statements are given. In these items a statement is made concerning some group of *persons*, or *events* or *institutions*, and the subject is to indicate how general these statements are. The words *All Most Many Few No* placed before each statement offer a choice of underlinings to indicate the degree to which the statement is believed to be true. In many psychological respects this test is similar to the preceding—it asks for a judgment on a statement. But whereas the former asks for the degree of conviction for or against the statement, the present asks for an estimate of the degree to which the statement can be said to be true. Those who think statistically will see that a man's belief of the degree of truth of a statement, if based on reasoning, would depend upon his estimate of the probabilities in the situation, and this leads back to the generalization test. However, opinions are determined by so many other factors than reasoning that the relationship between the two tests cannot be pushed too far. Examples are:

- All Most Many Few No Men of prominence in business to-day worked their way up from humble beginnings without money or influential friends to help them.
- All Most Many Few No Socialists are anxious to take away the money from the rich so they can have more for themselves.
- All Most Many Few No Poor men win important lawsuits against great corporations.

The other three tests, instead of getting at opinion directly, do so indirectly by testing the reasoning that one does in connection with the issue. Two of the tests, Form C and Form E, are based on the assumption that extreme, prejudiced attitudes in

either direction are either *based on* or *lead to rationalizations*. Form E is based on the supposition that persons who hold extreme opinions tend to *belittle* the arguments and evidence in favor of the opposite point of view and try to *stress* or *emphasize* the arguments and evidence in favor of their own stand. Working on this assumption, Watson raises an issue and then gives six arguments, three of which are commonly held to be in favor of the issue and three against it. Certain of these arguments are strong or good arguments, and others are weak or poor arguments. The directions require the subject to indicate which are strong arguments and which are weak arguments. Now it is assumed that if a person is unbiased, he will discriminate between the strong and weak arguments and mark them correctly; but if he is biased, all arguments on the side he favors will appear to him strong and all arguments on the other side will appear to him weak. The degree to which he favors one side or the other, and which side this is, is disclosed by the extent to which he errs in judging the arguments. Watson decided, because he did not succeed in getting arguments which competent people could agree were strong or weak, to mark an item biased only when there was unanimity in rating as strong all arguments in favor of one side of an issue and as weak all arguments on the opposite side.

Example:

- | | | |
|-----------|------|---|
| | | Is Socialism desirable in the United States to-day? |
| 1. Strong | Weak | It would give to all the people control of the natural resources now in the hands of a few. |
| 2. Strong | Weak | It would give over a great deal of control to men who are not refined or cultured, sometimes not respectable, and hence would be undesirable. |
| 3. Strong | Weak | Government enterprise has not proved as efficient in many ways as has private business. |
| 4. Strong | Weak | Socialism is desirable because it would take away money from those who have a great deal and would divide it up among the rest of the people. |
| 5. Strong | Weak | Socialists are undesirable radicals and extremists. |
| 6. Strong | Weak | The old parties have become so corrupt that the country should turn to Socialism. |

220 Diagnosing Personality and Conduct

In this example credit was given as being an economic radical if numbers 1, 4, and 6 were rated strong.

TABLE 34
RATINGS OF ARGUMENTS
(from Watson)

<i>Argument number</i>	<i>Side of issue</i>	<i>Test rating</i>	<i>Per cent of agreement of judges</i>
1	Yes	Strong	88%
2	No	Weak	83
3	No	Strong	100
4	Yes	Weak	92
5	No	Weak	92
6	Yes	Weak	95

In Form C, "Inference Test," certain facts pertaining to a situation are given. Following this brief description are certain conclusions or generalizations which one might draw from the facts. Those taking the test are instructed to check the conclusions which *fairly* follow from the facts as given, being careful not to assume anything else than the evidence given in the statement. They are also told, "You are not to consider whether the conclusions are right or true in themselves, but only whether they are rightly inferred from the facts given in the statement. You may check as many as you believe to be perfectly sure and certain. Do not check any merely probable inferences." Since most of the situations are described very briefly, no very important or extreme conclusions can logically be drawn. Accordingly any one who marks a conclusion extreme in either direction may be said to be biased or prejudiced.

Example:

Statistics show that, in the United States, of 100 men starting out at an age of 25, at the end of 40 years one will be wealthy and 54 will be dependent upon relatives or charity for support.

1. ☐ The present social order cheats the many for the benefit of the few.
2. ☐ The average young man, under present conditions, cannot count on being wealthy at the age of 65.
3. ☐ Most men are shiftless, lazy, or extravagant; otherwise they would not need to be dependent.
4. ☐ The one man is living upon luxuries ground out of the bones of the mass of common people.
5. ☐ Some day the workers will rise in revolt.
6. ☐ None of these conclusions can fairly be drawn.

In this example a person is given credit as tending toward an economic radical if he marks 1, 4, or 5.

Form D, "Moral Judgment Test," is a clever test based on the supposition that persons tend to be more biased about contemporary, immediate, personal affairs than about historical, distant, impersonal affairs. The test consists of pairs of described situations, one about a historical or distant event and another about a contemporary or immediate event. Statements are then given which approve or disapprove or in some way pass judgment on the story, one of which is to be marked if agreed with.

Example:

VI. During the latter part of the Great War and the years immediately following, officials of the United States government, suspecting certain organizations of being disloyal, broke into the headquarters of radical groups without legal warrant. Searching the premises, they found publications of questionable character and confiscated them, and collected evidence enough to convict some of the ring-leaders.

1. ☐ Such action was quite right, where radicals were suspected of opposing the Constitution.
2. ☐ The matter is indifferent, not to be called either right or wrong.
3. ☐ Such action on the part of the government officials was wrong, or at least unwise.

XII. A group of detectives were hired to investigate the activities of a great business corporation suspected of using some underhanded methods. Without any legal warrant, some detectives and some dissatisfied workmen of the company broke into the office, found the books, and proved that the corporation had been dishonest.

1. ☐ Since they got the evidence, what they did was all right.
2. ☐ The matter is indifferent.
3. ☐ To get evidence in such a manner was wrong, or at least not very wise.

If XII-1 was checked, together with either VI-2 or VI-3, the person taking the test showed a tendency to be biased in the direction of economic radicalism. The same tendency was shown if XII-2 and VI-3 were marked.

These tests were constructed on the hypothesis that one's expressed attitude is based on or leads to *rationalization*. In the Argument Test this was supposed to be shown by the tendency

222 Diagnosing Personality and Conduct

to overlook the logic in an argument; in the Inference Test, by the tendency to imagine things in a meagerly described situation that do not necessarily exist in the situation; and in the Moral Judgment Test, by the lack of consistency when the situation is impersonal and when it grows warm to the person's prejudice. The most doubtful points are whether a person's expressed attitude is necessarily the result of rationalization, and whether rationalization is the inevitable concomitant of an extreme position. It is conceivable that a person might hold an extreme opinion or make an extreme choice without necessarily sacrificing the rectitude of his logical processes. On the other hand it is an uncertain assumption that attitudes are based on or even very closely related to reasoning processes. Perhaps the test does not assume this, for it may be that the possessor of an unreasoned attitude resorts to rationalization only when hard pressed with the outcome of his reasoning processes in the test.

It may be that these tests are good tests of attitudes because in them a person shows the side on which he stands regardless of what his reasoning processes may be. Are directly expressed attitudes more valid than attitudes expressed under a subterfuge? If it is true that conduct is best measured by tests with a purpose so disguised that the person shows unconsciously his hand while he thinks he is doing something else, then these can be said to be attitude tests. But better results may come from a direct attack. The matter needs extended experimental study.

In later work with religious education tests, Watson shows his ingenuity in using new-type objective test methods for measuring attitude. In general the method is to get the pupil to answer questions so as to express choices, preferences, or opinions. Examples of the various types of tests in addition to those already described are (28, p. 17):

a. A brief description of an ethical problem situation with alternative solutions, one of which is to be marked best.

Example: If you make a mistake and put a nickel for a penny in the slot machine,

- a. Put in four slugs to even it up.
- b. Call up the company and tell them about it.
- c. Smash the thing and get your nickel.
- d. Report it to the police.
- e. Do nothing.

b. A brief description of a problem situation with alternative solutions to be ranked in order of desirability.

Example: A boy's parents ask him to help at home. He may
 Help when he feels like it.
 Choose certain jobs and see that these are always done.
 Promise to help but give younger brother candy to do the work instead.

c. A brief description of a problem situation with alternative solutions, one to be rated as the best solution, another as the worst solution.

Example: A bunch of boys are going to a dance in a town some distance away on a school night. One boy's parents think he should stay home and study.
 He may
 Say: "Oh, I'm going anyhow," and go.
 Sneak out and go.
 Think it out, fairly, for himself.
 Say, "Oh, all right," stay home and sulk.

In his "Orient and Occident" Watson (29) has a questionnaire to study attitudes toward Oriental problems. He divides this questionnaire into two sections, one to determine "How You Feel," the other to determine "What You Think." An example of the first is:

Directions: Read each word listed in capital letters in the column below and think quickly how you feel about it. Notice your own immediate reaction to it before you read further. Then read the words or phrases suggested about it, noticing which comes nearest to agreeing with your own reaction. Write the number of that word or phrase in the parenthesis in the right hand margin. If none seems just right, choose the one which comes nearest to expressing your feeling. If several appeal to you, choose the one truest to your first quick response. Do not try to reason out the logically best one.

1. JAPANESE (1) Alert and progressive; (2) Untrustworthy; (3) Courteous; (4) Ingenious; (5) Conceited; (6) Politically ambitious. ()

An example of the second is:

Directions: Please indicate your opinion about each of the statements below by drawing a circle around the letter or letters

224 Diagnosing Personality and Conduct

in the margin which expresses your judgment. This is what the letters mean.

T = True (absolutely)
 PT = Probably or Partly True
 D = In Doubt, Divided, Open Question
 PF = Probably or Partly False
 F = False (absolutely).

If you do not know enough about any item to express an opinion about it, cross it out.

T PT D PF F 1. Japan's growing population problem cannot be solved unless white peoples allow free Japanese immigration into their countries.

The distinction between thinking and feeling seems a little difficult to draw in this connection. It may be true that it requires somewhat more reasoning ability to extract the meaning from the statement than to catch the significance of the isolated phrases. But it would seem as though the real process of making a choice was about the same in the two instances—one based neither on thinking, nor on feeling, but on habitual responses of acceptance or aversion to one or another element in the situation presented.

Hornell Hart has done pioneering work in the use of the questionnaire in testing social attitudes and interests. His method (8, p. 9) consists essentially in presenting to groups of individuals, who are known to differ widely in recognized directions with respect to their *socialization*, stimuli so selected that it might be possible, on the basis of the observed differences in the reactions of these contrasted individuals to the stimuli, to construct tests and devise methods of scoring which would indicate the probable degrees of socialization of other individuals whose attitudes it is desired to determine."

The method of marking the test, novel at the time when Hart first used it, is indicated by the following sample:

Sample List X

Stub your toe	+	⊖
Tie your shoe	+	—
Come to despair	+	⊖
<u>Turn a page</u>	+	—
Fill your pen	+	—
<u>Receive \$1,000</u>	⊕	—

Mail a letter	+	-
<u>Be loved</u>	\oplus	-
<u>Be insulted</u>	+	\ominus
<u>Brush your hair</u>	+	-
Have a bath	\oplus	-
<u>Be seasick</u>	+	\ominus

The instructions to the subject are to put a circle around the plus sign \oplus after each thing that he is sure he likes, and around the minus sign \ominus after each thing he is sure he dislikes. Things about which he does not care one way or the other are to be skipped. A line is to be drawn under each of the five things that he feels most strongly about and a double line is to be drawn under the one thing in the list about which he feels most strongly of all.

This method of indicating responses does not work entirely satisfactorily. In using this procedure in school, one can never tell whether a pupil is omitting an item through carelessness or neglect or because he is actually neutral to that item. Consequently, it is better to require that every item be answered and then give a neutral symbol which may be encircled to indicate neutrality. Another possible defect is the limitation placed on the expression of extreme feeling by requiring that five items be underlined once and one of these doubly underlined. Greater freedom of expression for individual differences in feeling may be obtained, if desired, by including symbols to be encircled for indicating degree of feeling.

Four lists in Hart's questionnaire deal with things which people like (or dislike) to be, to do, or to have happen. Lists 5 and 6 contain lists of things to read or to study. Tests 7 and 8 contain reforms which the subject may favor or disfavor.

This "test" of Social Attitude and Interests was given on the one hand to a group of thirty-three men, who were taken as "leaders of social progress," and on the other hand to a miscellaneous group of 154, including thieves, employed boys, grade and high-school boys, business men, college business men, successful business men, junior "medics," and other college men. The responses of these two groups were tabulated separately. When they were compared, it was found that the two groups gave markedly different reactions on ninety-three of the 149 stimulus

words. Of course there is the possibility that factors other than degree of socialization such as amount of schooling or native ability may influence the results, but Hart considers this possibility and dismisses it. Hart concludes that (8, p. 37) "the differences in reactions are such as to indicate that the men in the highly socialized group either are, or believe themselves to be, or think it desirable to appear, markedly more interested in international, economic, criminal and social justice, far more interested in discovering truth and having it spread abroad freely, more interested in intellectual and ethical aspects of religion, but less interested in creeds and forms, far less interested in conventionality and social approval, decidedly less sentimental and dominated by sympathy and immediate personal bonds, and much more indifferent to light reading, to certain aspects of personal comfort, to business success, and in general to trivial and selfish interests than the other men tested are, or believe themselves to be, or think it desirable to appear."

Shuttleworth (18) has used a similar technique to measure the character and environmental factors other than intelligence involved in scholastic success in college. He used an instrument which he calls an "assayer," which consists of a questionnaire of attitude and interests, a self-rating scale and a questionnaire about home and school background. The self-rating section has been omitted in a more recent revision called "Student Information Blank." The results of this study of Shuttleworth's are not clear inasmuch as he did not use a criterion which sharply differentiated between scholastic achievement and intelligence. His correlations would seem to show that he was not successful in extracting items which could be used as measures of achievement apart from ability.

The present writer (21) used a modification of the Hart technique to measure students' interests and attitudes. In its revised form the questionnaire contained the following lists of items:

- List 1. Things one would like to own.
 2. Occupations.
 3. Activities.
 4. Places to go.
 5. Magazines.
 6. School activities.
 7. Things to be judged right or wrong.

Before each item the three symbols + : - were placed, one of which was to be encircled to indicate liking, neutrality, or disliking. Pupils were instructed to omit no items. The 100 most significant items were selected, using as a criterion the studiousness index (see p. 333) or teachers' ratings on school industry. Papers of 296 high school seniors in three schools were scored, using these 100 items. If an item was answered + or - in the direction of studiousness it was given a credit of 2; if the item was answered neutrally, it was given a credit of 1; and if it was answered in the opposite direction to that of studiousness it was given a credit of zero.

The self- r (fifty items against fifty items) was .732 ($N = 80$) for one school, and .695 ($N = 213$) in another school. Using the Spearman-Brown formula, the reliability of the whole questionnaire of 100 items is found to be .846 and .820 respectively.

The correlations obtained in one school were as follows:

TABLE 35
INTERCORRELATIONS OF SCHOOL MARKS, RATINGS FOR STUDIOUSNESS, STUDIOUSNESS INDEX, AND STUDIOUSNESS QUESTIONNAIRE
(from Symonds, 21)

	Marks	Ratings	Studiousness Index	Studiousness Questionnaire
Terman Group Test of Mental Ability339	.017		-.337
Average of school marks..		.556		.121
Teachers' rating of studiousness475	.232
Studiousness index397

In another school the correlations found were:

* teacher rating on studiousness—studiousness questionnaire	.373
* average mark—studiousness questionnaire	.252

Assuming the use of a technique of questioning similar to that devised by Hart, and his method of evaluating the responses by the use of contrasted groups, this type of questionnaire has great possibilities.

"There is no reason why questionnaires cannot be assembled to diagnose every phase of man's activities and interests, using a technique similar to that outlined above. With empirical methods of selecting and rejecting items, it should be possible to find

items which correlate significantly with every phase of man's life. In the measurement of conduct, investigators have been blocked because conduct does not leave behind a permanent record which may be studied and measured. This can be accomplished by great effort and expense by setting up special situations in which an objective record may be obtained as has been done by the Character Education Inquiry. There is always danger in these methods that a narrow or artificial sampling of the precise conduct to be studied may be tested or that extraneous variables, as for example motives, may vitiate the results. It should, however, be possible to find by trial and error methods verbal responses which will correlate with almost any conduct trend." (21, p. 167.)

Another distinct technique for measuring attitudes, and one which is still more analytic than any previously described, was developed by Allport and Hartman (1). This technique consists in getting a wide variety of opinions on some issue and then scaling these opinions from one extreme position to the other. This graduated scale of opinions may then be read over by any one whose attitude is being measured, and he may indicate which statement best expresses his own opinion. The sample of the scale chosen then becomes a measure of attitude.

In their original work Allport and Hartman chose seven concrete issues of current interest (1, p. 736), "dealing, respectively, with the League of Nations, the qualifications of President Coolidge, the distribution of wealth, the legislative control of the Supreme Court, prohibition, the Ku Klux Klan, and graft in politics. Sixty students, upperclassmen, were asked to write their personal views on the various phases of these questions. The resulting opinions on each issue were then carefully sifted and the distinct and relevant views were assembled. Keeping the issues separate, these views were presented on slips of paper and arranged independently by six judges, teachers of political science and psychologists, in order of their logical position in a scale ranging from one extreme on an issue in question to the opposite extreme." Later copies of the scales were distributed to students in another class, and each student was instructed to read the scale carefully, and to check the one statement in each of the seven issues which most nearly coincided with his or her own view. They were also given opportunity to designate the *certainty* of their opinion and the *intensity of interest or feeling* upon the question.

Thurstone (23, 24, 25), sensing the importance of Allport's work, has embarked on a program of developing Allport's basic idea and refining the statistical methods employed, which is aimed to eventuate in a series of scientifically determined and defensible scales.

One of Thurstone's scales is reproduced here:

*Scale of Attitude toward the Movies **

(The scale value of each statement is shown in parentheses following its serial number. The higher the scale value the more favorable the statement toward the movies. The statements have been arranged in random order.)

1. (1.5) The movies occupy time that should be spent in more wholesome recreation.
2. (1.3) I am tired of the movies; I have seen too many poor ones.
3. (4.5) The movies are the best civilizing device ever developed.
4. (0.2) Movies are the most important cause of crime.
5. (2.7) Movies are all right, but a few of them give the rest a bad name.
6. (2.6) I like to see movies once in a while, but they do disappoint you sometimes.
7. (2.9) I think the movies are fairly interesting.
8. (2.7) Movies are just a harmless pastime.
9. (1.7) The movies to me are just a way to kill time.
10. (4.0) The influence of the movies is decidedly for good.
11. (3.9) The movies are good, clean entertainment.
12. (3.9) Movies increase one's appreciation of beauty.
13. (1.7) I'd never miss the movies if we didn't have them.
14. (2.4) Sometimes I feel that the movies are desirable, and sometimes I doubt it.
15. (0.0) It is a sin to go to the movies.
16. (4.3) There would be very little progress without the movies.
17. (4.3) The movies are the most vital form of art to-day.
18. (3.6) A movie is the best entertainment that can be obtained cheaply.
19. (3.4) A movie once in a while is a good thing for everybody.
20. (3.4) The movies are one of the few things I can enjoy by myself.
21. (1.3) Going to the movies is a foolish way to spend your money.

* From Thurstone, L. L., "A Scale for Measuring Attitude Toward the Movies." *Journal of Educational Research*, 22: 93, 94 (Sept., 1930).

230 Diagnosing Personality and Conduct

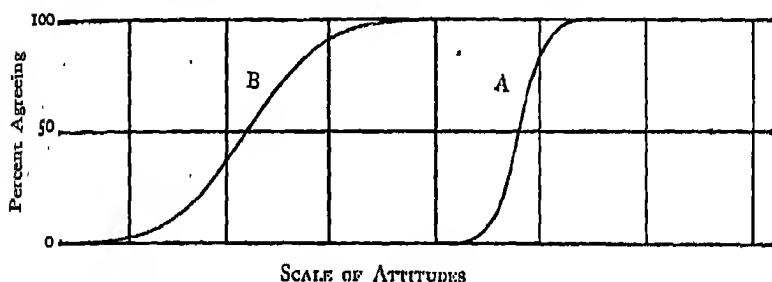
22. (1.1) Moving pictures bore me.
23. (0.6) As they now exist movies are wholly bad for children.
24. (0.6) Such a pernicious influence as the movies is bound to weaken the moral fiber of those who attend.
25. (0.3) As a protest against movies we should pledge ourselves never to attend them.
26. (0.1) The movies are the most important single influence for evil.
27. (4.7) The movies are the most powerful influence for good in American life.
28. (2.3) I would go to the movies more often if I were sure of finding something good.
29. (4.1) If I had my choice of anything I wanted to do, I would go to the movies.
30. (2.2) The pleasure people get from the movies just about balances the harm they do.
31. (2.0) I don't find much that is educational in the current films.
32. (1.9) The information that you obtain from the movies is of little value.
33. (1.0) Movies are a bad habit.
34. (3.3) I like the movies as they are because I go to be entertained, not educated.
35. (3.1) On the whole the movies are pretty decent.
36. (0.8) The movies are undermining respect for authority.
37. (2.7) I like to see other people enjoy the movies whether I enjoy them myself or not.
38. (0.3) The movies are to blame for the prevalence of sex offenses.
39. (4.4) The movies is one of the great educational institutions for common people.
40. (0.8) Young people are learning to smoke, drink, and pet from the movies.

In using the scale for placing a person's attitude, Thurstone prefers not merely to ask a person which sample on the scale most nearly coincides with his point of view, but would also have him indicate the range of opinions which he would be willing to endorse. The mean position of this range might not coincide with the step on the scale which he would state as best representing his position, but the mean might be a fairer measure of his opinion.

In selecting statements for inclusion in the scale Thurstone (24, 25) recognizes the following criteria: (a) The statements should be as brief as possible so as not to fatigue the reader who is asked to read the whole list. (b) The statements should be

such that they can be endorsed or rejected in accordance with their agreement or disagreement with the attitude of the reader. (c) Every statement should be such that acceptance or rejection of the statement does indicate something regarding the reader's attitude about the issue in question. (d) Double-barreled statements should be avoided. (e) Irrelevant statements should be avoided.

Thurstone gives three criteria by which items of the scale may be judged for validity. (a) The scale must transcend the group measured. In other words it should be a true scale for both pacifists and militarists; or for wets and dries. (b) The items must be unambiguous. This means that the ogive curve rises steeply from 0 per cent agreeing to 100 per cent agreeing as one proceeds up the scale of attitudes. In the diagram, item A is less



ambiguous than item B. (c) The items must be relevant. The criterion of relevance is to see the extent to which those in a group who endorse the item tend to endorse items higher up the scale and fail to endorse items lower down the scale.

Thurstone has performed some valuable analytical work in the method of scale construction from judgments preparatory to the actual work of making a set of attitudes scales.

Reliability

The reliability of the various attitude measures used has been meagerly determined. The extension of methods has outstripped their critical study.

Watson (30) has studied the reliability of his test both for total prejudice score and for the separate scores for the eighteen lines of bias. Results of .92, .89, and .96 were obtained for the

232 Diagnosing Personality and Conduct

self-correlation in three groups, half of the test against half. The reliability of the whole test may be placed at .96. The following table shows the reliability found in separate biases.

TABLE 36
RELIABILITY OF ATTITUDE QUESTIONNAIRES
(from Watson)

<i>Direction of Bias</i>	<i>Self-r (30 papers)</i>	<i>Self-r (61 papers)</i>	<i>Maximum number of points</i>	<i>Probable self-r on 200 such points (70 papers)</i>
1. Economic radicals74	.69	189	.70
2. Economic liberals39	.51	93	.69
3. Economic capitalists ..	.68	.59	192	.60
4. Persons fighting for a "social gospel"84	.69	188	.70
5. Persons interested in a "personal gospel"82	.84	143	.88
6. Religious fundamen- talists94	.79	182	.81
7. Religious modernists..	.81	.60	89	.77
8. Religious radicals97	.80	164	.83
9. Protestants64	.35	58	.65
10. Catholics51	.52	66	.77
11. Persons with high, strict moral standards.	.78	.69	107	.81
12. Persons with broad, loose moral standards.	.70	.46	94	.65
13. Bias against economic capitalism75	.67	188	.68
14. Bias against economic radicals54	.49	205	.49
15. Bias against religious fundamentalists40	.58	151	.65
16. Bias against religious radicals68	.84	151	.88
17. Bias against the unsci- entific46	.58	86	.76
18. Bias against the re- ligious85	.78	182	.80

The present writer (20) found a reliability of .67 for the 115 items on his social attitudes questionnaire. Thirty-two of the most significant items yielded a correlation of .62, which would be raised to .83 for 120 equally significant items. Zeleny (31) found a reliability of .893 and .894 for a sixty-eight item true-false opinion test. Harper (6) found that a true-false test of

social attitudes yielded reliability coefficients for one half of the test with the other half of .782, .751, and .817 in various groups of teachers. Jones (10) finds a correlation of .71 between the twenty-five odd and twenty-five even items of an opinion test with college students in which an opportunity is given for expressing an opinion as to five degrees of truth for each item.

Validity

Attitude questionnaires, as measures of opinion, are valid only to the extent that they agree with other indications of a person's choice or tendency to act. However, an expressed opinion is important on its own account, regardless of how well it agrees with the actual overt choice that might be made, or the tendency to act. In a democracy, voting, public speaking, and other forms of expression help determine the policies of the state. There is every reason for taking the results of an attitude questionnaire at face value as an expression of opinion, particularly when no immediate issue is at stake. But as measures of underlying tendencies to act they should be checked against other indications of choice before being accepted as valid.

Allport's work shows that different issues may result in different types of opinion distributions. He obtained some normal, some skewed, and some bimodal distributions within the college group which he studied. There was a distinct cleavage of opinion concerning the League of Nations, and an overwhelming sentiment for prohibition with the group tested. Rice, on the other hand, believes that great bodies of social opinion as expressed in elections form a normal distribution. A really satisfactory answer to this question must await technical work on the evenness of scaling. Even the possible significance of Allport's bimodal distributions must not be overlooked, although certainly in attitude questionnaires made up of a number of items designed to measure more general attitude tendencies, the distribution tends to approximate normality.

Allport finds that there is a distinct relationship between the tendency to select expressions of opinion at either extreme end of a scale and *certainty* of opinion. "Certainty of the kind here exhibited is not an index of objective truth, but an accompaniment of increasing distortion of truth through narrowed emphasis

234 Diagnosing Personality and Conduct

upon one phase. The man thinks he is right because he feels strongly." Allport found that intensity of feeling and certainty tend to go together, and also that there are more men than women at the extremes at both ends of the scale.

Allport studied his group by means of self-ratings and various types of questionnaires. He concluded that the radicals and reactionnaires, that is, those taking extreme positions, are more like each other than the conservatives or middle-of-the-roaders in self-ratings of emotionality, rapidity of movement, self-reliance, overestimation of mental ability, and lack of agreement with the conventional moral code. In the attitude questionnaires it was found that radicals and reactionnaires share one another's attitudes on diverse questions more fully than the conservative shares the attitudes of either. Again, without complete evidence, Allport surmises that the radical resembles the *extrovert* and the reactionary resembles the *introvert*, a conclusion which remains to be demonstrated.

Watson (30) considered the validity of his tests on a number of grounds. He found the correlation of each sub-test with the total as follows:

TABLE 37
CORRELATIONS BETWEEN GROSS SCORES FOR EACH TEST FORM AND GROSS
SCORES FOR TOTAL TEST
(from Watson)

<i>Test form</i>	<i>Correlation</i>
A	.68
B	.94
C	.79
D	.58
E	.56
F	.53

Watson says,

"It seems fair, therefore, to conclude that each form of this test is as closely related to the purpose of the test as a whole, as are the various items in an intelligence test to the score of the test as a whole; and that the forms are more closely related in these tests of fair-mindedness than is commonly the case in testing various forms of arithmetical ability, reading ability, writing ability and the like."

Watson uses other means of validating his test. He compared the score of certain individuals selected by their group as being

fair-minded with the score of others in the groups, especially with other groups with the reputation of being decidedly prejudiced, and found that the selected individuals actually did rate high in fair-mindedness on the test. Case studies tended to reveal the accuracy of the diagnosis afforded by a study of the test results. Comparison of the scores of groups which by reputation are fair-minded or prejudiced indicated that the tests substantiated popular impression.

The present writer (20) found a correlation between his social attitude questionnaire and social information of .28, and intelligence of .28. Jones (10) reports a correlation of less than .20 between his measure of radicalism and intelligence. Watson finds that his test correlates $-.005$ with the Thorndike-McCall Reading Test and $-.08$ with IQ. Apparently these measures of attitude have little relationship to ability.

Symonds found a definite sex difference, the boys tested tending to be more liberal. Harper found that "male graduate educators were found to occupy a position on the scale slightly less conservative than the position of female graduate educators." Jones reports little sex difference.

Jones found that the college course had slight effect on the opinion of college students. Seniors were more conservative than freshmen in economic affairs, but more liberal in religious matters. Symonds found no change in social attitude from the eighth grade through the college senior year. On the other hand Harper reports a correlation of .521 between extent of education above the eighth grade and scores on his test of radicalism. He found that a group which pursued a graduate course in education which stimulated freedom of discussion in the atmosphere of liberalism raised its score 14.4 points on his test of seventy-one items during the year. Watson also reports that in a Y. M. C. A. conference and in certain classes on race problems there was a distinct increase in scores on fair-mindedness. This would point to the fact that training in attitudes is very specific, that ordinary school instruction does not have much influence on social attitudes, but that carefully planned activities and discussions are capable of modifying attitudes and prejudices in the particular field worked in.

Harper reports no relationship between attitude and religious affiliation or the political party espoused.

Watson (29) in his study of attitudes toward the Orient finds that the results on his questionnaire show negligible relationship with geographical sections of the country and amount of education. There is evidence that racial attitudes are related to economic level, religious denomination, the reading of liberal periodicals, and having Oriental friends. He also finds a marked relationship between amount of reading and a liberal attitude and between information and attitude. These relationships interlock, and until further work is done it is impossible to tell which are fundamental and which are merely subsidiary.

Do these questionnaires measure attitude? Can attitude be measured? Bain * insists that we have no surety of a high relationship between verbal and adjustment behavior, especially when tabooed or ill-organized behavior is the subject of investigation. Thurstone's (23) interesting discussion is pertinent here.

There comes to mind the uncertainty of using an opinion as an index of attitude. The man may be a liar. If he is not intentionally misrepresenting his real attitude on a disputed question, he may nevertheless modify the expression of it for reasons of courtesy, especially in those situations in which frank expression of attitude may not be well received. This has led to the suggestion that a man's action is a safer index of his attitude than what he says. But his actions may also be distortions of his attitude. A politician extends friendship and hospitality in overt action while hiding an attitude that he expresses more truthfully to an intimate friend. Neither his opinions nor his overt acts constitute in any sense an infallible guide to the subjective inclinations and preferences that constitute his attitude. Therefore we must remain content to use opinions, or other forms of action, merely as indices of attitude. It must be recognized that there is a discrepancy, some error of measurement as it were, between the opinion or overt action that we use as an index and the attitude that we infer from such an index." **

But some outer expression must be taken as the sign or symbol of the inner indication or choice. The objective data show evidence that a person's verbal expression of his opinion may be taken as a fair index of his attitude. It is usually safer in measuring conduct to disguise what one is doing. If you tell a

* Bain, R., "Theory and Measurement of Attitudes and Opinions," *Psychological Bulletin*, 27: 357-379 (1930).

** Thurstone, L. L., "Attitudes Can Be Measured." Reprinted by permission from the *American Journal of Sociology*, Volume XXXIII, No. 4.

person directly that you are measuring his prejudice he will at once be on his guard. So those attitudes tests or questionnaires ought to be most successful which make a person divulge his preference or choice when he is apparently intent on some other activity. However, the direct question or ballot also seems to yield satisfactory measures of attitude when there is no immediate and personal issue at stake.

REFERENCES

1. ALLPORT, F. H., and HARTMAN, D. A., "Measurement and Motivation of a Typical Opinion in a Certain Group," *American Political Science Review*, 19: 735-760 (1925).
2. BOGARDUS, E. S., "Analyzing Changes in Public Opinion," *Journal of Applied Sociology*, 9: 372-381 (1924-1925).
3. ———, "Measuring Social Distance," *Journal of Applied Sociology*, 9: 299-308 (1924-1925).
4. ———, "Social Distance and Its Origins," *Journal of Applied Sociology*, 9: 216-226 (1924-1925).
5. BUSCH, HENRY M., "Race Attitudes of Children," *Religious Education*, 21: 277-281 (June, 1926).
6. HARPER, M. H., *Social Beliefs and Attitudes of American Educators*, Teachers College Contributions to Education, No. 294 (1927), 91 pp.
7. HART, H. N., "Research Possibilities with a Socialization Test," *Journal of Applied Sociology*, 7: 163-166 (1923).
8. ———, *A Test of Social Attitudes and Interests*, University of Iowa Studies in Child Welfare, 2, No. 4 (1923), 40 pp.
9. ———, "Can We Measure Results in Teaching Social Science?" Proceedings of the Nineteenth Annual Meeting of the American Sociological Society, Chicago, 1924.
10. JONES, E. S., "The Opinions of College Students," *Journal of Applied Psychology*, 10: 427-436 (Dec., 1926).
11. MANRY, J. C., *World Citizenship: An Ethical Theory of Higher Education*, doctor's dissertation, University of Iowa, 1923.
12. MOORE, H. T., "Innate Factors in Radicalism and Conservatism," *Journal of Abnormal and Social Psychology*, 20: 234-244 (1925).
13. NEUMANN, GEORGE B., *A Study of International Attitudes of High School Students*, Teachers College Contributions to Education, No. 239 (1926), 120 pp.
14. RICE, S. A., "Differential Changes of Political Preference under Campaign Stimulation," *Journal of Abnormal and Social Psychology*, 21: 297-303 (Oct.-Dec., 1926).

15. RICE, S. A., "Some Applications of Statistical Method to Political Science Research," *American Political Science Review*, 20: 313-329 (May, 1926).
16. ———, "The Political Vote as a Frequency Distribution of Opinion," *Journal of the American Statistical Association*, 19: 70-75 (March, 1924).
17. SHUTTLEWORTH, F. K., "A New Method of Measuring Character Traits," *School and Society*, 19: 679-682 (1924).
18. ———, *The Measurement of the Character and Environmental Factors Other Than Intelligence Involved in Scholastic Success*, University of Iowa Studies in Character Education, 1, No. 3.
19. STURGES, H. A., *Studies of the Dynamics of Attitude*, unpublished. (Washburn College.)
20. SYMONDS, P. M., "A Social Attitudes Questionnaire," *Journal of Educational Psychology*, 16: 316-322 (1925).
21. ———, "A Studiousness Questionnaire," *Journal of Educational Psychology*, 19: 152-167 (1928).
22. TENDLER, A. D., "Measurable Factors of Adjustment in Juvenile Delinquents," *Jewish Social Service Quarterly*, 1: 1-31.
23. THURSTONE, L. L., "Attitudes Can Be Measured," *The American Journal of Sociology*, 33: 529-554 (Jan., 1928).
24. THURSTONE, L. L., and CHAVE, E. J., *The Measurement of Attitude* (University of Chicago Press, 1929).
25. THURSTONE, L. L., "The Measurement of Opinion," *Journal of Abnormal and Social Psychology*, 22: 415-430 (1928).
26. WASHBURN, M. F., et al., "The Moore Tests of Radical and Conservative Temperament," *American Journal of Psychology*, 38: 449-452 (1927).
27. WARSON, G. B., "Does World Mindedness Depend upon Good Will or Information, upon Character or Intelligence?" *Religious Education*, 21: 188-191 (April, 1926).
28. ———, "Experiments with Religious Education Tests," Program Papers, No. 5 (Association Press, 1926, 50 pp.)
29. ———, *Orient and Occident*, confidential report to the American Group of the Institute of Pacific Relations, 1927. See also a summary of this report in *Religious Education*, 24: 322-328 (1929).
30. ———, *The Measurement of Fair-Mindedness*, Teachers College Contributions to Education, No. 176 (1925), 97 pp.
31. ZELENY, L. D., "A Measure of Social Opinions of Students," *Journal of Applied Sociology*, 11: 56-64 (Sept.-Oct., 1926).

Chapter VII

INTEREST QUESTIONNAIRES

IN using interest questionnaires for guidance, two questions of fundamental importance must be considered. One concerns the *permanency of interest*, the other, the *relation of interest and ability*. Unless interests have some degree of permanence, the determination of interests will have little prognostic significance. If interests are unstable, fickle, subject to every passing whim, instead of representing deep and underlying trends, they have little prophetic value, they are of only theoretical importance, and they have little bearing on the welfare of any one.

An investigation of how interest and ability are related is necessary in order to help determine the relative importance of each in guidance. If interest and ability are closely correlated, one may be used in place of the other, and a single determination of ability is all that is necessary for giving adequate advice; if, however, interest and ability are not closely related, separate determinations of each are needed so that one may supplement the other.

Permanence of Interest

Studies of the permanence of interest have been made by Thorndike (7, 8), Willet (9, 10), Crathorne (1), Franklin (2, 3), and Fryer (4). Thorndike (8) had college juniors rank in interest the following seven activities as they remembered them in elementary school: mathematics, history, literature, science, music, drawing and other handwork. Similar rankings were obtained for what the subjects remembered their order of interest to be during the high school period and also for their present order of interest in college. The correlation of order of interest in elementary school and in college was on the average over .60. Thorndike realizes that recollection of previous interests is not altogether trustworthy, but offers these considerations in its defense (8, p. 453): "I do not believe that such tendencies to read present

interests into the past and to leave the order reported for one period unchanged so far as possible, are very strong, there being a contrary tendency to remember and look for differences. On the whole, I should expect the effect of the large chance errors in *lowering* the estimate of permanence to nearly or quite counteract whatever balance of prejudice there may be in favor of similarity of interests or projection of present conditions into the past." Thorndike concludes (8, p. 456): "These facts unanimously witness to the importance of early interests. They are shown to be far from futile and evanescent. It would indeed be hard to find any feature of a human being which was a much more permanent fact of his nature than his relative degrees of interest in different lines of thought and action."

Willet (9, 10) obtained answers on three successive occasions, at one-year intervals, to the same questions on choice of school subjects and future occupations from 488 junior high school pupils. He concludes: "It appears that permanence of interests both in school subjects and in future vocations is decidedly lacking for the majority of these 488 pupils." A technique of asking pupils their present interests and repeating the questions after an interval is probably safer than the technique used by Thorndike, in which memory is depended on. There is, however, a possibility that Willet used a classification of occupations which was so fine that a pupil might not be credited with repeating the same answer merely because he changed the *name* of the occupation without really indicating a change in the fundamental character of his interests. Willet believes this factor is not important, for whereas he found that 121 changes were from first to second choice, ninety-nine pupils interchanged first or second best-liked subject with the most-disliked subject. He also believes from a study of the reasons given for disliking subjects that pupils in stating interest in a subject are influenced by their liking or disliking of the teacher.

Crathorne (1), using a technique similar to Thorndike's, found that of those pupils who professed definite leanings toward some particular occupation on entering high school, almost exactly one half claimed a radical change of interests before entering college.

Franklin (2, 3) has made the most thorough and extensive study of the permanence of vocational interests by repeatedly obtaining answers to the same questions regarding present interests

from the same children. He reports, after carrying his work through three years, that (3, p. 440), "the vocational interests of pupils show a high degree of permanence during the junior high school period. After three years, three children out of four still adhere to their original *type* of vocational preference and three out of five still cling to the actual choice which they originally expressed. . . . The high degree of constancy of the percentage of permanency over so long a period of time indicates that the interests expressed by entering junior high school pupils are significant and worthy of consideration."

Fryer, in summing up his own work, in which he used the technique of Thorndike, gives the following table which represents his own judgment of the matter.

TABLE 38
PROBABLE PERMANENCY OF VOCATIONAL AMBITION BETWEEN VARIOUS
DEVELOPMENTAL PERIODS

(from Fryer, 4, p. 477)

	<i>Expressed in percentages</i>	<i>Expressed in correlation coefficients</i>
Grammar-school and high school	75	+ .4
High school and college	75	+ .5
Grammar-school and college	50	+ .1
Grammar-school, high school, and college.....	40

Fryer's conclusion after surveying his findings is (4, p. 478): "Vocational interests cannot be used as exact guides to future vocational interests; they are only slightly suggestive of such future interests."

The upshot of this conflicting evidence is that statements by young people of *precise* preferences of either subjects of study or occupations are of little value inasmuch as they are unstable and subject to change. On the other hand, statements in terms of broad categories, whether of subjects of study or occupational groups, represent rather deep and underlying trends and can be depended on to possess considerable predictive significance. In infancy interests are capable of taking almost any turn, being dependent on only the most general factors such as keenness of the senses, intelligence, muscular power, etc. On the other hand, as we shall see, in adulthood interests have become rather firmly established so that only infrequently and with great difficulty does a man make a radical change in his occupation or even avocation.

Between these two extreme periods the interests are being molded. At first the broad trends are laid down, and interest becomes inclined toward books, people, mechanical objects, animals, plants, sport, etc. Gradually even these become specialized, so that normally the period of development is also the period for the specialization of interests.

At the college level these interests have on the whole become so set that interest questionnaires have great prognostic significance, as the work of Moore (37), Ream (40), Freyd (30), Strong (49), and Cowdery (25, 26) indicates. On the other hand in the junior high school, because differentiation of interests is not nearly so marked, much less dependence can be placed on definite statements of interest, notwithstanding that the broad underlying trends of interest are probably pretty well established at that time. We cannot deny that cleverly constructed questionnaires of a number of items designed to differentiate between these underlying interest trends may be of much value and significance for guidance even in the junior high school period.

Interests and Abilities

The relation between interests and abilities has been studied by Thorndike (23, 24), Hartman and Dashiell (17), Bridges and Dollinger (11), and Fryer (13, 14). Thorndike (23), as an addition to the class experiments described in the previous section on permanency of interests, also obtained data at a later meeting of the same classes showing rankings in ability with respect to the seven categories which he used. He states that reports of relative interest at ages eleven to fourteen have a median correlation of .66 with relative ability at twenty-one or later. He concludes, "A person's relative interests are an extraordinarily accurate symptom of his relative capacities."

Bridges and Dollinger (11) criticized Thorndike's experiment on the grounds that a person's ranking of his ability would influence or conversely be influenced by his ranking of interest. Using success in college courses as a measure of ability, these investigators find that the correlation of ability and interest ranges from .22 to .28. They state, "The conclusion of practical importance for vocational guidance is that both must somehow be evaluated separately."

Thorndike (22), commenting on the work of Bridges and Dol-

linger, gives as his belief that the method which these writers used underrated the true relationship. Furthermore, he points out that the range of interests embodied in the set of school subjects which any one student carries is probably much less than the range of interest in the seven subjects which he (Thorndike) used in his previous experiments. Thorndike maintains after reviewing the evidence that a fair figure for the true relationship of interest and ability is .70.

Hartman and Dashiell (17), in a laboratory experiment with paper and pencil tests probably covering a narrow range of interest, found a negligible relationship between success on the tests and interest in them.

Fryer (16) has studied the relationship between the intelligence of a pupil and the intelligence level of the occupation which he states to be his choice. The relationship is not at all close. The following table presents Fryer's conclusion based on his findings.

TABLE 39

PROBABLE RELATION OF INTELLIGENCE AND INTELLIGENCE REQUIREMENT OF VOCATIONAL AMBITION AT VARIOUS DEVELOPMENTAL PERIODS

(from Fryer, 4, p. 489)

<i>Developmental periods</i>	<i>Per cent likely to approximate correct choice</i>	<i>Correlation coefficient expressing relationship</i>
In grammar-school (later years).....	30	+0.1
In high school (other than at graduation).....	40	+0.2
In college (other than at graduation).....	35	+0.1
In vocational school (requiring high school graduation)	75	+0.5
In vocations (seeking vocation guidance, ages 20 to 30 years).....	50	+0.4

Fryer concludes, "It would appear that there is some relation between intelligence and ability to select a suitable vocation. . . . Vocational interests should be considered as one of many factors to be used in vocational guidance."

In *Genetic Studies of Genius*, Volume I, Terman presents data showing how gifted children (with high IQ's) differ in interests from average children. In scholastic, occupational, play, and reading interests there are noticeable specific differences between bright and average children. A coefficient of correlation of .41 expresses the average relationship between the order of preferences of studies as stated by the children and order of quality of work in

the subjects as rated by the teachers. Since the correlations which go to make up this average were found on control and gifted groups separately, they undoubtedly suffer from restricted range.

Kelley (18) in his *Educational Guidance* tried out tests of ability and interest for predicting success in various subjects. He gives the correlation of tests of ability and interest for predicting mathematics, English, and history as .35, .34, and .33 respectively.

Recent studies have discussed more in detail the relation of interest in school studies to academic achievement. There is a marked tendency toward good achievement in subjects that students profess to like and toward low achievement in subjects they dislike. Langlie concludes (21, p. 248), "the results indicate that there is a relationship between statements of interest and grades obtained in single courses; and the relationship is probably significant enough to be of value to an adviser or personnel man. There is a tendency to obtain one's best grades in those courses (in college) which were liked in secondary schools, and to get lower grades in those courses which were disliked in secondary schools."

But recent attempts to use interests for the prediction of achievement have been disappointing. King (19) presents such results as the following:

TABLE 40
CORRELATION OF INTEREST IN SCHOOL SUBJECTS AND ACHIEVEMENT
(from King)

	<i>Achievement in English</i>	<i>Achievement in mathematics</i>	<i>Achievement in science</i>
Interest in English.....	.143	-.136	-.193
Interest in mathematics.....	-.105	.342	-.110
Interest in science.....	-.029	-.260	.187
Composite interest165	.361	.276

In summary we must conclude that the relation between ability and interest is distinct but not close, and that the whole problem is a somewhat confused one. Too much care cannot be taken to define what is meant by *ability*. If *ability* is used synonymously with *intelligence*, it is evident from Fryer's data and from common observation that the range of interests for persons of the same intelligence may be very wide. If interest and ability in a single activity are compared, we find the relationship positive but

low. The relationship becomes close only when a comparison is made between ability and interest in rather fundamental or basic activities where there is a possibility for wide differences in both interest and ability.

Our genetic scheme used in discussing the permanence of interests applies here. In infancy, when neither interest nor ability has developed, the relationship is near zero. In adulthood, when ability in various activities has reached its probable (though not possible) maximum and interests have become specialized, the relationship is fairly close. Between these two periods the relationship is increasing. In childhood, when the broad interest trends are taking form and abilities are entering upon differentiation, the relationship is low. The young child who has no skill in swimming probably has no special interest in swimming. As he increases his skill, his interest keeps pace. The expert in any line whose *ability* is *recognized* is almost sure to parallel his ability with interest. The two increase together. In the junior high school period, however, the relationship is low enough so that one forms a useful supplement to the other for the purposes of guidance.

The conclusion is that properly and skilfully devised questionnaires that differentiate between rather fundamental interest trends ought to prove valuable devices for the guidance of adolescents.

Measurement of Interests

A fruitful line of attack upon the measurement of interests was developed under the direction of C. S. Yoakum in the now dissolved Bureau of Personnel Research at the Carnegie Institute of Technology. One of the first studies in this group, by B. V. Moore, attempted to find factors that would differentiate the abilities and interests of salesmen, designers, and production executives who were graduates from the engineering school. Moore was interested in finding out for each graduating engineer in what line of work he would function most efficiently and satisfactorily. Of all the tests he used, he found that a questionnaire of occupational preferences showed the greatest differentiation. His method was the empirical one of trying out the questionnaire on apprentice engineers and determining the responses for each item. Moore describes his scoring method as follows (37, p. 45):

"A stencil in the form of a cardboard with perforations or slots allowing only the engineering type of occupations to be visible is

246 Diagnosing Personality and Conduct

placed over the list of occupations which have been checked. The number of plus marks is counted and recorded in the margin. The number of minus marks is also recorded. Then this stencil is removed and another stencil is placed over the list, allowing only the sales type of occupation to be visible; and then the number of plus and minus marks is recorded. The number of *plus* marks before engineering occupations is added to the number of *minus* marks before sales occupations in order to get the number of checks in favor of engineering occupations. The number of *minus* marks before engineering occupations is added to the number of *plus* marks before sales type of occupations to get the number of checks in favor of sales occupations. Finally the number of marks in favor of sales occupations is divided by the total number of check marks to get the percentage of marks in favor of sales occupations."

Ream (40), in a study of life insurance salesmen, used a more extensive blank. Ream's contribution to the technique was his selection of items that showed only a difference which satisfied a statistical criterion of reliability. Otherwise each item was weighted *one* in making his composite.

Freyd (30) used two questionnaires devised in the Bureau of Personnel Research which profited from the experience of Moore and Ream. One was a list of seventy-two occupations designed to sample all types of activities. Each occupation was followed by L ? and D, one of which symbols the subject answering the questionnaire was to encircle according to his wish to indicate liking, neutrality, or disliking of the occupation. The second list contained 129 items which a person might like or dislike. These included a series of physical attributes of people; a series of mental attributes of people; and a series of miscellaneous items. Each of these items was followed by five symbols as shown to permit the recording of five degrees of liking-disliking.

Fat men	L!	L	?	D	D!
Fat women	L!	L	?	D	D!

Freyd, interested in finding items which would differentiate between the "mechanically minded" and the "socially minded," used the following method:

1. For each group tables were made showing the frequency with which each symbol was encircled.

Interest Questionnaires

247

Item	Group	Response					Total
		L!	L	?	D	D!	
Fat men	Mechanical	2	14	19	6	2	43
	Social	2	3	16	6	3	30

2. Select those symbols which seem to show significant differences in the proportions of the two groups encircling them. In the above item the responses 14 and 3 for the symbol L seem to be significantly different.

3. Determine the differences in proportion.

$$\frac{14}{43} = .33 \quad \frac{3}{30} = .10$$

4. Determine the standard error of the difference in proportion by the formula.

$$e^2 = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

In the above example,

$$e^2 = \frac{.33 \times .67}{43} + \frac{.10 \times .90}{30}$$

$$= .0081$$

$$e = .09$$

$$\frac{d}{e} = \frac{.23}{.09} = 2.56$$

5. List the cases in which the difference in proportion is at least twice as great as its standard error.

6. Determine which items will enter into the score in a positive way and which in a negative way. One group must be chosen to represent the positive direction. Each item is given a value of +1 or -1; that is, all items are weighted equally.

7. The total score for any person will be the algebraic sum of the positive and negative values attached to the significant items which he checks.

Since there are many symbols which are not included in the scoring scheme, it will often mean that a person will escape marking the scorable symbol merely by chance. Freyd, confronted with this criticism, answered it as follows (30, p. 87): (a) Symbols were not chosen for the scoring key unless a "fair" proportion of the number of persons answering the item checked that symbol. (b) Those who use some other symbol than that in the scoring

248 Diagnosing Personality and Conduct

key really are given an intermediate score (o) between the two key scores (+1 and -1) and are hence not really ignored. (c) "We may assume that the factors which operate to cause one to go into an occupation are very diverse, and that any two or three are sufficient to place him in that occupation. Then as long as an individual falls into compartments showing significant differences in two or more tests or answers to questionnaires, he shows evidence of having been influenced by those two factors to place him in the occupation in which he finds himself." These arguments are not wholly convincing, especially since Freyd admits that one or two items can influence the score.

The Bureau of Personnel Research Interest Analysis Blank has been taken up and rigorously studied by Cowdery and Strong at Stanford University. Their modified "Interest Report Blank" included eighty-four occupations, seventy-eight types of people, thirty-four sports and amusements, six kinds of pets, thirteen representative kinds of reading, twenty-three miscellaneous activities, and twenty-five school subjects. Adults require twenty to thirty minutes to fill out the blank.

Cowdery (25) first gave the questionnaire to three groups of "carefully selected" professional people: thirty-four doctors, thirty-seven engineers, and thirty-four lawyers. Scoring keys were worked out for each group separately. In all previous work with this questionnaire each item has been weighted *one*. Through the influence of T. L. Kelley, Cowdery adopted the following plan of weighting scores. This weight, called *b*, is obtained from the formula

$$b = \frac{\phi}{(1 - \phi^2)\sigma}$$

where ϕ is the coefficient of correlation from the formula

$$\phi = \frac{ad - bc}{\sqrt{(a + c)(b + d)(a + b)(c + d)}}$$

This latter formula demands a fourfold table and necessitates a dichotomization of the L ? D series of possible responses. We are not told how this dichotomization was accomplished, but probably the frequency of replies to one symbol was used against the sum of replies to the other two. These *b*'s yielded weights for each item in the Medical Interest Scale from 0 to 13, with a total pos-

sible score for all 263 items of 920. Cowdery found that when the members of all three professions were scored on the Medical Interest Scale there was very little overlapping between doctors and members of the other professions. Similar results were found from scales designed to measure interest in the other professions. Using the same set of questions, scoring keys were developed which gave highly differentiated responses corresponding with remarkable closeness to the interests of different professional groups.

Strong (43), using the same Interest Analysis Blank, studied the responses from eighteen different occupational groups. He discovered a method of weighting items which was simpler than the Kelley-Cowdery method, yet yielded equally good results. Strong (43, p. 201) found "that there exists a nearly straight-line relationship between the *b* scores and the difference between the percentage of the profession in question and the percentages of all the professions. This straight-line relationship holds very well if either, or both, of the percentages are between 8 per cent and 92 per cent inclusive. Within these limits differences of 0 to 2 are weighted 0; 3 to 7 are weighted 1; 8 to 11, 2; 12 to 15, 3; 16 to 20, 4; etc. If either of the percentages is between 0 and 2 or 98 and 100 inclusive, the differences are weighted as follows: 0, 0; 1, 1; 2 to 3, 2; 4 to 6, 3; 7 to 10, 4; 11 to 14, 5; 15 to 18, 6; etc. An intermediate set of weights are necessary if either percentage is between 3 and 7 or 93 and 97. Use of this table possibly introduces a slight error in some cases but the great saving in time makes possible the calculation of many more data."

Strong's (43) method, then, for determining a scoring key for personnel managers, for example, is as follows:

1. The percentages of personnel managers who like, are indifferent to, and dislike each of the 263 items are obtained.

<i>Item</i>	<i>L</i>	<i>I</i>	<i>D</i>
Actor	49%	38%	13%

2. The percentages of men from eighteen occupations who like, are indifferent to and dislike each of the items were obtained.

<i>Item</i>	<i>L</i>	<i>I</i>	<i>D</i>
Actor	38%	35%	27%

3. The differences between these two sets of figures are obtained.

<i>Item</i>	<i>L</i>	<i>P</i>	<i>D</i>
Actor	+11	+3	-14

4. These figures are replaced by smaller figures according to the scheme described on page 249 as follows:

<i>Item</i>	<i>L</i>	<i>P</i>	<i>D</i>
Actor	+2	+1	-3

5. An individual is scored by the use of this key.

Strong (44) has recently (1927) thoroughly revised and extended this questionnaire, now calling it "Vocational Interest Blank." The experience with the questionnaire to date leads Strong to maintain that "members of each occupational group have characteristic likes and dislikes which distinguish them from other occupational groups." Such evidence as is at hand indicates that "these vocational interests are present in college freshmen and are not materially altered by technical training and subsequent professional experience."

The possibility that this same type of interest questionnaire might prove to be valuable in helping boys to discover their interests when a first choice of curriculum has to be made in the high school led Garretson (32) to assemble a set of items for a tryout in three New York City high schools—the De Witt Clinton High School (an academic high school), the High School of Commerce, and the Brooklyn Technical High School. His original questionnaire has been revised and at present consists of nine sections as follows:

TABLE 41
GARRETSON'S INTEREST QUESTIONNAIRE* (SECTIONS AND NUMBERS OF ITEMS)

<i>Section</i>	<i>Number of items</i>
1. Occupations	80
2. Activities	30
3. School subjects	20
4. Job activities	20
5. a. School paper	15
b. Football team	18
c. Student activities	18
6. Prominent men	12
7. Things to own	26
8. Magazines	24
Total	263

* Published by the Bureau of Publications, Teachers College, 1930.

In the scoring for Garretson's revised questionnaire weights of only +, 0, and — are to be used for each response according as it reliably indicates preference for, neutrality toward, or dislike of, any particular curriculum, because Garretson found in the course of his work that this simple method yields results as satisfactory as a more elaborate weighting system. Plus items are given a weight of 2; neutral items, 1; and negative items, 0. In scoring, the number of positive items is counted and multiplied by 2. The number of negative items is counted, added to the number of positive items, and the total subtracted from 263 to give the number of neutral items, which, since they are one each, are added to the value of the positive items for the total score.

In printing the questionnaire, the three possible responses to each item are printed in a column $\begin{smallmatrix} L \\ I \\ D \end{smallmatrix}$ so as to make the scoring procedure simpler.

TABLE 42
DISTRIBUTIONS FOR THREE SCHOOLS IN QUESTIONNAIRE, SCORED FOR
COMMERCIAL INTEREST *

Score	<i>Brooklyn De Witt Clinton</i>		
	<i>Technical High School</i>	<i>High School (Academic)</i>	<i>High School of Commerce</i>
460-479.....	5
440-459.....	19
420-439.....	..	1	23
400-419.....	1	9	45
380-399.....	6	20	53
360-379.....	8	33	55
340-359.....	15	52	69
320-339.....	49	69	65
300-319.....	52	60	47
280-299.....	73	89	29
260-279.....	83	64	27
240-259.....	86	54	11
220-239.....	85	23	6
200-219.....	52	16	2
180-199.....	31	10	..
160-179.....	14	3	..
140-159.....	1	..	1
120-139.....
100-119.....
80-99.....
60-79.....
Total.....	556	503	457
Mean.....	264.49	300.21	350.63
σ	47.34	49.59	55.03

* From Symonds, P. M., *Tests and Interest Questionnaires in the Guidance of High School Boys* (Teachers College, Bureau of Publications, 1930).

Reliability

Hubbard (35) studied the reliability of the original Freyd blank and found it to be low and unsatisfactory as follows:

TABLE 43
RELIABILITY OF FREYD INTEREST BLANK
(from Hubbard)

<i>Time between tests</i>	<i>Men</i>		<i>Women</i>	
	<i>r</i>	<i>N</i>	<i>r</i>	<i>N</i>
Six weeks52	(285)	.47	(313)
One year64	(156)	.49	(193)

Perhaps this low reliability may be explained in part by the scoring scheme, which utilized only a small fraction of the possible responses that might differentiate.

However, when the results are portrayed so as to show the amount of shifts from one division of the scale to another, there appears a considerable amount of stability, particularly among those who obtain extreme ratings.

TABLE 44
SHOWING CHANGES IN SCORE FROM ORIGINAL TEST (FREYD'S INTEREST ANALYSIS
BLANK) TO RETEST
(from Hubbard, 35, p. 621)

Number of cases	PER CENT OF CHANGE		
	156	285	
	<i>Year interval</i>	<i>Six weeks interval</i>	
Of those who were — 2 or less	56	58	remained — 2 or less
Of those who were — 2 or less	26	14	moved to between — 1 and + 1
Of those who were — 2 or less	18	28	moved up to + 2 or more
	100	100	
Of those who were + 2 or more	67	63	remained + 2 or more
Of those who were + 2 or more	25	25	moved to between + 1 and — 1
Of those who were + 2 or more	8	11	moved to — 2 or less
	100	100	

Cowdery (25) reports the following coefficients of reliability which were obtained from random halves of the questions, corrected by the Spearman-Brown formula.

TABLE 45
RELIABILITY COEFFICIENTS OF VOCATIONAL INTEREST BLANK
(from Cowdery, 25, p. 138)

	<i>Legal scale</i>	<i>Engineering scale</i>	<i>Medical scale</i>	<i>Average</i>
Experienced men905	.802	.775	.827
Graduate students846	.831	.778	.818
Upper division university students.	.813	.786	.794	.798
Lower division university students.	.782	.661	.685	.709
High school senior boys.....	.842	.844	.697	.798

These are fairly respectable reliability coefficients. Further work must be done to harmonize these findings with those of Hubbard. Garretson reports reliability coefficients obtained by correlating odd responses with even responses of .911 for technical preference, .861 for commercial preferences, and .756 for academic preference, which are raised to .953, .955, and .861 by the Spearman-Brown formula for the whole questionnaire.

Validity

The Carnegie Institute of Technology questionnaire has been tried out only in connection with vocational differentiation and has been found to yield remarkably promising results. Strong (44) has the records from eighteen occupations and is able to determine (a) the scores which members of any one profession made on scales for other occupations and (b) the scores which members of different occupations made on the scale of any one profession. Strong has used three letters A, B, and C to describe the position of a man on any scale. A means a score attained by 75 per cent of the criterion group, B means a score made by the lowest 25 per cent of the men for whom the scale was developed, and C means a score lower than that made by any man in the given occupation.

Strong (47) later found that members of many occupational groups rate as *executives* to a surprising degree; the overlapping in the case of the executive interest scale is three to four times that obtained from the other seven occupational interest scales.

TABLE 46

PERCENTAGES OF MEN IN VARIOUS OCCUPATIONS WHO RATE *A* AND *B* IN THE INTEREST OF CERTIFIED PUBLIC ACCOUNTANTS, ENGINEERS, AND PERSONNEL MANAGERS

(from Strong, 47, p. 233)

	Scale for certified public accountants		Scale for engineers		Scale for personnel managers	
	<i>A</i>	<i>B</i>	<i>A</i>	<i>B</i>	<i>A</i>	<i>B</i>
Certified public accountant...	74	25	2	28	3	31
Banker	20	39	2	31	0	12
Office worker	2	39	7	25	2	24
Lawyer	5	23	5	25	7	40
Engineer	2	21	75	25	3	31
Personnel manager	6	13	6	33	75	25
Author	2	16	4	18	2	10
School-teacher	2	15	8	29	4	31
Life insurance salesman	4	12	0	29	5	30
Advertising man	0	10	6	38
Doctor	0	8	9	42
Minister	0	0	0	10	2	34
Artist	0	0	4	29	0	0
Line executive	7	32
Department store salesman...	0	14

TABLE 47

CORRELATIONS BETWEEN SCORES OBTAINED BY CERTIFIED PUBLIC ACCOUNTANTS, ENGINEERS, AND PERSONNEL MANAGERS IN THEIR OWN AND OTHER OCCUPATIONAL INTERESTS

(from Strong, 44, p. 236)

<i>Scale</i>	<i>Certified public accountant</i>	<i>Engineer</i>	<i>Personnel manager</i>
Certified public accountant	1.00	-.21	.10
Line executive31	.37	.15
Lawyer19	-.36	.47
Banker16	-.09	-.39
Advertising man13	.08	.30
Engineer13	1.00	-.36
Office worker12	-.26	-.26
Author06	.06	.14
Artist03	-.01	-.23
Life insurance salesman01	-.08	.31
Personnel manager00	-.30	1.00
Minister	-.14	-.52	-.10
School-teacher	-.25	-.51	.00
Doctor	-.30	+.21	..
Department store salesman	-.04

Strong says that possibly among various occupations there are to be found a large percentage of potential executives.

Cowdery (25) has demonstrated that scales developed from the interest questionnaire correlate about zero with intelligence scores. "Significant relations with intelligence test scores are lacking; a slight positive relation with Thorndike test score was noted in the case of the respective interest ratings of law and engineering students."

Cowdery reports a correlation of .335 between engineering interest scores and university grades in engineering subjects. He also found that "freshmen and sophomores planning to be engineers, juniors and seniors in engineering school, engineers in graduate work, engineers with less than five years' practical experience score approximately the same on the same interest test. The same holds true with respect to physicians and lawyers." This evidence is presumptive that those preferences are not transient affairs but represent rather permanent turns in the individual interests. It also suggests that such interests do not arise from training, but rather tend to select people for various vocations.

Garretson (32) found that his questionnaire differentiates in a remarkable way between the interests of one curriculum group and another. Biserial r 's between the distribution of scores in one curriculum and another are given in the following table. These relationships were found on different groups from those from which the scoring keys were obtained.

TABLE 48
RELIABILITY COEFFICIENTS OF GARRETSON INTEREST QUESTIONNAIRE FOR HIGH SCHOOL STUDENTS

Questionnaire Scoring Key	Garretson's Data			Symonds' Data		
	Commer- cial against other two	Technical against other two	Academic against other two	Technical- Academic	Technical- Commer- cial	Academic- Commer- cial
Commercial....	.727	-.433	-.809	-.544
Technical.....868730	.706	.197
Academic.....560	-.637	-.274	.433

The relationships as expressed by these biserial r 's are higher than are commonly found between intelligence tests and later measures of achievement. In other words, this questionnaire predicts considerably more accurately the curriculum a boy will

256 Diagnosing Personality and Conduct

choose than an intelligence test will predict his success in that curriculum.

This approach to the measurement of interests promises to be very valuable to both educational and vocational guidance. Part of the success of the method lies in its application of the sampling theory. It is well recognized that a single statement of a boy's choice of occupation or other interest is comparatively unreliable. A week later the boy may have changed his mind. But the cumulative evidence from a hundred or so questions indicates reliably the direction of interest along one or more broad general lines. Interest questionnaires seem to be promising, not because they will answer questions as to choice of specific curricula or occupations, but because they point out certain broad trends of interest.

REFERENCES

Permanence of Interest

1. CRATHORNE, A. R., "Change of Mind between High School and College as to Life Work," *School and Society*, 11: 28-30 (1920).
2. FRANKLIN, E. E., *The Permanence of the Interests of Junior High School Pupils* (The Johns Hopkins Press, 1924), 63 pp.
3. ———, "The Permanence of Vocational Interests after Three Years," *School and Society*, 23: 438-440 (April 3, 1926).
4. FRYER, D., "The Significance of Interests for Vocational Success," *Mental Hygiene*, 8: 466-505 (1924).
5. MACKAYE, D. L., "The Fixation of Vocational Interest," *American Journal of Sociology*, 33: 353-370 (Nov., 1927).
6. PROCTOR, W., *Psychological Tests and Guidance of High School Pupils* (Public School Publishing Company, 1923).
7. THORNDIKE, E. L., "Early Interests: Their Permanence and Relation to Abilities," *School and Society*, 5: 178-179 (1917).
8. ———, "The Permanence of Interests and Their Relation to Abilities," *Popular Science Monthly*, 81: 449-456 (1912).
9. WILLET, G. W., "Permanence of Pupil Interests," *School and Society*, 7: 325-330 (March 16, 1918).
10. ———, "Permanence of Pupil Interests," *School and Society*, 9: 334-338, 365-368 (1919).

Relation of Interest and Ability

11. BRIDGES, J. W., and DOLLINGER, W. M., "The Correlation between Interests and Abilities in College Courses," *Psychological Review*, 27: 308-314 (1920).

12. COMMINS, P. S., and SHANK, T. B., "The Relation of Interest to Ability in Secondary Subjects," *Elementary School Journal*, 27: 768-771 (June, 1927).
13. FRYER, D., "Intelligence and Interest in Vocational Adjustment," *Pedagogical Seminary and Journal of Genetic Psychology*, 30: 127-151 (1923).
14. ———, "Interest and Ability in Educational Guidance. A Statistical Study," *Journal of Educational Research* 16: 27-39 (1927).
15. ———, "Predicting Abilities from Interests," *Journal of Applied Psychology*, 11: 212-225 (1927).
16. ———, "The Significance of Interests for Vocational Purposes," *Mental Hygiene*, 8: 466-505 (1924); also *Vocational Education Monographs*, No. 3: 35-45 (1924).
17. HARTMAN, R., and DASHIELL, J. F., "An Experiment to Determine the Relation of Interests to Abilities," *Psychol. Bull.*, 16: 259-262 (1919).
18. KELLEY, T. L., *Educational Guidance*, Teachers College Contributions to Education, No. 71 (1914), 116 pp.
19. KING, L. H., *Mental and Interest Tests—Their Evaluation and Comparative Effectiveness as Factors of Progress in Secondary Education*, unpublished doctor's dissertation, Columbia University, 1929.
20. KITSON, H. D., "Interest as a Criterion in Vocational Guidance," *Educational Review*, 53: 349-356 (1916).
21. LANGLEY, T. A., "Interests and Scholastic Proficiency," *Personnel Journal*, 9: 246-250 (1930).
22. THORNDIKE, E. L., "The Correlation between Interests and Abilities in College Courses," *Psychological Review*, 28: 374-376 (1921).
23. ———, "Early Interests: Their Permanence and Relation to Abilities," *School and Society*, 5: 178-179 (1917).
24. ———, "The Permanence of Interests and Their Relation to Abilities," *Popular Science Monthly*, 81: 449-456 (1912).

Interest Questionnaires

25. COWDERY, K. M., "Measurement of Professional Attitudes," *Journal of Personnel Research*, 5: 131-141 (Aug., 1926).
26. ———, "The Interest Inventory in College Vocational Guidance," *The Psychological Clinic*, 19: 59-62 (1930).
27. CRAIG, D. R., "The Preference-Interest Questionnaire in Selecting Retail Saleswomen," *Journal of Personnel Research*, 3: 366-374 (1925).
28. FREYD, M., "A Method for the Study of Vocational Interests," *Journal of Applied Psychology*, 6: 243-254 (1922).

258 Diagnosing Personality and Conduct

29. FREYD, M., "The Measurement of Interests in Vocational Selection," *Journal of Personnel Research*, 1: 319-328 (1922).
30. ———, *Personalities of the Socially and Mechanically Minded*, Psychological Monographs, Vol. XXXIII, 4, No. 151 (1924), 85 pp.
31. FRYER, D., *The Measurement of Interests* (Henry Holt and Company, 1931).
32. GARRETSON, O. K., *Relationships between Expressed Preferences and Curricular Abilities of Ninth Grade Boys*, Teachers College Contributions to Education, No. 386 (1930).
33. HUBBARD, R. M., "Interests Studied Quantitatively," *Journal of Personnel Research*, 4: 365-378 (Jan.-Feb., 1926).
34. ———, "A Measurement of Mechanical Interests," *Journal of Genetic Psychology*, 35: 229-254 (1928).
35. ———, "The Reliability of Freyd's Interest Blank," *Journal of Educational Psychology*, 17: 617-624 (Dec., 1926).
36. KORNHAUSER, A. W., "Results from a Quantitative Questionnaire on Likes and Dislikes Used with a Group of College Freshmen," *Journal of Applied Psychology*, 11: 85-94 (Feb., 1927).
37. MOORE, B. V., *Personnel Selection of Graduate Engineers*, Psychological Monographs, Vol. XXX, 5, No. 138 (1921).
38. ———, "A Tested Method of Using Tests for Vocational Guidance," *School and Society*, 18: 761-784 (1923).
39. POFFENBERGER, A. T., "The Measurement of Character and Interest," in *Applied Psychology* (D. Appleton and Company, 1927), 586 pp.
40. REAM, M. J., *Ability to Sell* (The Williams and Williams Company, 1924), 64 pp.
41. Research Committee of the National Vocational Guidance Association upon the Measurement of Interest, "The Outlook in the Measurement of Interest," *Personnel Journal*, 9: 176-183 (1930).
42. RULON, P. J., and ARDEN, W., "A Scoring Technique for Tests Having Multiple Item Weightings," *Personnel Journal*, 9: 235-241 (1930).
43. STRONG, E. K., JR., "The Interest Test for Personnel Managers," *Journal of Personnel Research*, 5: 194-203 (1926-1927).
44. ———, "Differentiation of Certified Public Accountants from Other Occupational Groups," *Journal of Educational Psychology*, 18: 227-238 (April, 1927).
45. ———, "Interest Analysis of Personnel Managers," *Journal of Personnel Research*, 5: 235-242 (Oct., 1926).
46. ———, "Procedure for Scoring an Interest Test," *The Psychological Clinic*, 19: 63-72 (1930).

Interest Questionnaires

259

47. STRONG, E. K., JR., "Vocational Guidance of Executives," *Journal of Applied Psychology*, 11: 331-347 (Oct., 1927).
48. ———, "Vocational Guidance of Engineers," *Industrial Psychology*, 2: 291-298 (June, 1927).
49. ———, "Vocational Interest Test," *The Educational Record*, 8: 107-121 (April, 1927).
50. SYMONDS, P. M., *Tests and Interest Questionnaires in the Guidance of High School Boys* (Teachers College Bureau of Publications, 1930).

Chapter VIII

TESTS OF CONDUCT KNOWLEDGE AND JUDGMENT

IN a discussion of methods of measuring conduct, tests of knowledge and judgment certainly should not be neglected. Just what relation they have to conduct will appear in the sequel. Without anticipating our discussion, we may say that it seems only natural that men should have turned to tests of knowledge and reasoning for the diagnosis of conduct. Even the popular belief that action follows knowledge, that we reason out beforehand the course of action we are to pursue, gives warrant enough to investigators to experiment with tests of this type. This belief permeates our institutions, our educational theories and practices, and indeed all our relations with our fellows. Courts of criminal law decide responsibility for an act on the basis of the defendant's ability to discriminate between right and wrong.

In the following discussion different types of tests will be described, some of the better tests in detail. Following this, correlations will be reported which show exactly what relations the abilities shown by the tests have to measures of conduct and intelligence. Finally, the significance of the tests and the bearing of knowledge on conduct will be discussed.

Tests of the Vocabulary of Conduct

Fundamental to tests of knowledge and discrimination is ability to understand the vocabulary of the field of conduct under consideration. There are few vocabulary tests in special fields, and none is available which covers conduct in the fields of health, the use of language, the handling of property, etc. Good vocabulary tests have been constructed in the social-ethical field by Kohs (34) and Schwesinger (40), and a special test of slang by Schwesinger (39) is available. Kohs' vocabulary test (number 4 of his battery called "Ethical Discrimination Test" *) is in the multiple-response form and consists of forty-five items; of which samples follow:

* Published in 1922 and sold by C. H. Stoelting Co., Chicago.

Tests of Conduct Knowledge and Judgment 261

- (a) bad means clean, wrong, both, good.
 (b) revenge means sheep, bear with, treat, get even.
 (c) guilt means offense, guillotine, sword, golden.

By far the best piece of work in this field is the social-ethical vocabulary study of Schwesinger, conducted under the direction of the Character Education Inquiry. This special vocabulary comprises the words commonly used to describe situations involving human relations (*joke, company*), terms used in deciding moral issues (*illegal, villain*), adjectives which denote modifications of character (*bashful, recalcitrant*), abstract nouns indicative of states of mind and character traits (*uncertainty, snobbishness*), and verbs indicating behavior of human beings towards each other (*scoffing, pitying*). A few "shady" words consisting of slang and professional crook terms were included. Thorndike's "Tests of Word Knowledge" were accepted as a model, and 1,000 vocabulary items were constructed. Seven hundred of these are for testing words which occur in the first 10,000 of common usage as defined by the "Thorndike Word List." These 1,000 words were broken up into five groups of 200 each. The five tests thus constructed were given to school-children, and an elaborate item analysis followed. Three hundred most "symptomatic" items were separated into two forms of 150 words each, made equal in difficulty on the basis of the earlier testing. These tests, when given to children in grades five through eight, showed very high reliability. Seventy-five items in Form A correlated with the other seventy-five .966. The same correlation in Form B was .961. Corrected so as to apply to the whole test the correlation becomes .98. Discussions of the significance of the test will be given later.

Schwesinger (39) had also constructed earlier a test of the vocabulary of slang which she hoped would have some significance with relation to character and would help separate delinquents from non-delinquents. The test has four sections as follows (39, p. 254 f.):

I. Definitions—25 items.

Samples

- | | |
|--------------|---|
| Simp | means dumbbell, simpleton, sink, scab. |
| Ivory Dome | means cupola, bonehead, elephant's tusk, foolish. |
| Iron Men | means crusaders, dollars, bones, rubes. |
| Sheeney | means Silky, Kike, Jips, Jew. |
| To be Stewed | means drunk, cooked, tanked, batty. |

262 Diagnosing Personality and Conduct

II. Phrases—48 items.

Samples

Let George Do It means
 Pass the buck
 Avoid accepting responsibility
 Let George try if he wants to
 To get away with it
 He Has A Screw Loose means
 He needs to be mended
 He isn't quite sane
 He is a hunk o' cheese
 He is dippy

III. Same-Opposites—50 items.

Samples

cold	hot	Same	Opposite
dragged out	all-in	Same	Opposite
stall	jostler	Same	Opposite
razz	tease	Same	Opposite
hand-out	food	Same	Opposite

IV. Classification—18 lines, 124 items.

Samples

Write the letter S under every word that is used in Smoking.
 Write the letter G under every word that is used in Gambling or Cards.
 Write the letter D under every word that has something to do with Drinking.
 Write the letter T under every word connected with some form of Thieving.
 Write the letter F under every word that has to do with Fighting.
 Write the letter A under every word that has to do with Getting Arrested.
 Write the letter W under every word that has to do with Woman.
 Write the letter C under every word that has to do with Craps.

1. nicked, moll, four-flushing, jimmying, sand-bag, whopper, salt-creek.
2. deal, dead-head, switched, booze, skirt, Little Joe, moll-buzzer.
3. slacker, nicotine, pulled, kick, scalawag, dive, rough-house, flivver.

Tests of Knowledge of Conduct

Tests have been made of the knowledge of conduct in the field of health by Gates and Strang (22) and in the field of manners by Orr (27). In other phases of conduct, knowledge tests have not been developed. This is probably indicative of the fact that direct knowledge of desirable conduct is not a subject of instruction. In most phases of conduct society is probably not quite sure enough of the validity of its mores to codify them and test the knowledge of the codes.

The "Gates-Strang Test in Health Knowledge" is a very fine contribution to the growing collection of standardized tests. It is based on a very extensive and thorough analysis and inventory of facts and principles of hygiene found in twenty selected courses in health offered in rural and city schools in America, and in fourteen of the most widely used textbooks on health. These facts and principles were judged for their validity by experts. Seven hundred and forty-four exercises in the multiple-choice form were constructed and given to a large number of pupils in elementary schools, high schools, and colleges. Of these 224 were eliminated for such reasons, according to the authors, as, "They were essentially duplicates of others; no scientifically correct answer could be made at the present time; the results from testing showed ambiguous, misleading or confusing statements; the experts thought the idea too trivial or too technical to include."

The result of the research is a list of 520 items classified under topical headings and arranged from easiest to hardest in each group, with difficulty values appended. From this list a teacher may select items to form a test of any length, level of difficulty, or range of difficulty, or on any topic or combination of topics. For those who wish printed tests, short forms of sixty-four items selected from various topics and arranged according to difficulty have been prepared.*

Samples of the exercises are given below:

1. We should have fresh air
 - ☒ all the time
 - ☐ in the daytime but not at night
 - ☐ at night but not during the daytime
 - ☐ especially in summer
 - ☐ when we begin to get a headache.

* Published by the Teachers' College Bureau of Publications.

264 Diagnosing Personality and Conduct

2. Boys and girls should brush their teeth
 - twice a year
 - once a month
 - twice a day
 - once a week
 - twice a week.
3. Keeping school desks neat and floor free from papers should be done by
 - the teacher
 - the janitor
 - the principal
 - the pupils
 - the parents.

Miss C. I. Orr, working for the Character Education Inquiry, has made a "Test of Good Manners." This is an information test of the current standards of courtesy and good manners, using three types of tests—the true-false, the multiple-choice and the single-answer. Samples of the items follow (27, p. 199):

True-False

In helping yourself to sugar always use your own spoon.

True False

When yawning, make no attempt to suppress it by covering the mouth.

True False

Multiple-Choice

Approval of a program may be shown by

1. Stamping feet
2. Clapping
3. Whistling

Yes-No

Should a man tip his hat to a strange lady when picking up an article which she has dropped?

Yes No

If a door is closed, is it necessary to knock before entering a friend's room?

Yes No

Tests of Biblical knowledge. It seems a far cry to include tests on the Bible in a discussion of the diagnosis of conduct, and one must understand the history of the movement to comprehend the reason for the relationship between the Bible and conduct. The reasoning involved has never been very direct. On the one hand, since the Protestant Reformation made religion depend not

Tests of Conduct Knowledge and Judgment 265

on the authority of the church but on the authority of the Word as it is printed in the Bible, it became extremely important that Protestants should know how to read the Bible as the guide to salvation. The zeal shown by some early missionaries in teaching the heathen to read was for the sole purpose of enabling them to read the Holy Scriptures. On the other hand, Christianity is interpreted to mean a way of living, *mores*, a set of customs. In church school instruction, both the Bible and the Christian virtues and standard of living are taught. It is little wonder that some relationship between the two should be sought. This relationship has been rationalized by asserting that the Bible is filled with all kinds of social and ethical relationships and that the teachings of Jesus are the bases for the Christian standard of living. Accordingly, in some quarters, ethical instruction and Bible instruction are synonymous. In view of this far-reaching belief, it certainly behooves us to include tests of Bible knowledge in our review so that the exact relation between Bible knowledge and conduct may be determined.

The first of these "Biblical Knowledge Tests" appearing in 1920 was prepared by M. T. Whitley* and was designated the Old Testament, Series A. This test was intended for children nine years old and upward who are above the primary department of the Sunday-school. There are five tests.

<i>Test no.</i>	<i>Title</i>	<i>Form of item</i>	<i>Number of items</i>
I.	Relationships and Location by Books	Multiple-choice	30
II.	Source of Quotations	Single-answer matching	10
III.	Order of Bible Books	Two alternatives multiple-choice	15
IV.	History Facts	Single-answer	40
V.	Completed Quotations	Completion	11

This test may be criticized as testing a very superficial knowledge of the Bible because it is mainly concerned with its mechanical make-up, "historical" facts, and recognition of quotations. It fails entirely to test knowledge of the teachings, point of view, or

* Published by M. T. Whitley, Teachers College, Columbia University, New York City.

266 Diagnosing Personality and Conduct

development of the Bible. It represents the traditional objectives of Sunday-school teaching.

Dr. Whitley has since published a New Testament test of the Bible.

Another test also published in 1920 was by J. T. Giles (23, 24). This was a simple test consisting of three parts in the true-false form. The first twenty-five questions cover information on the Old Testament; the second twenty-five questions cover information on the New Testament; and the last twenty-five are questions involving ethical judgment.

Giles' test was revised in 1924 by W. L. Hanson, who turned all items into the best-answer or multiple-choice form and made some other minor changes. Two forms are available.

A test superior to those so far mentioned was constructed by S. R. Laycock—the "Laycock Test of Biblical Information."* The test comes in seven parts as follows:

<i>No. of test</i>	<i>Subject of test</i>	<i>Type of item</i>	<i>No. of items</i>
I.	General information	Multiple-choice	15
II.	Knowledge of what certain passages contain	Multiple-choice	8
III.	General information	True-false	30
IV.	Knowledge of what certain passages teach	Multiple-choice	8
V.	General information	Multiple-choice	15
VI.	Miscellaneous knowledge items	Multiple-choice	8
VII.	General information	Multiple-choice	16
			<hr/> 100

The distinction between the parts is not clear-cut. Probably there is an advantage in splitting up the test into separate subtests in order to break the monotony of a long, unbroken string of items for the younger pupils. But the separate parts seem to contain no special diagnostic significance.

An "Advanced Bible Knowledge Test" by G. B. Watson and* Eliot Porter is published as one of the series of "Character and Personality Tests" of the Association Press. It is composed as follows:

* Published in 1922 and for sale by the University of Alberta Bookstore.

Tests of Conduct Knowledge and Judgment 267

<i>No. of part</i>	<i>Subject of test</i>	<i>Type of item</i>	<i>No. of items</i>
Ia.	Old Testament	Multiple-choice	17
Ib.	New Testament	Multiple-choice	28
II.	Brief outline of the con- tents of the Bible	Completion	50
			—
			95

This test involves a more highly developed critical ability than the simpler information tests previously mentioned.

Besides these tests which deal specifically with Bible information there are other tests dealing with religious ideas and judgment.

The "Chassell Test of Religious Ideas Involving the Ranking of Selected Answers" by C. F. Chassell-Cooper and L. M. Chassell-Toops (14) provides a check on the pupil's religious ideas. The following questions are asked:

1. What is the purpose of the Old Testament?
2. What is the purpose of the New Testament?
3. How do you think of God?
4. How would you describe God?
5. What does God do?
6. How do you think of Jesus?
7. Why should we pray?

Following each of these questions is a list of several possible answers, and the pupil is instructed to choose the three best answers, marking them 1, 2, and 3 according to first, second, and third choice.

There are also tests of the comprehension of proverbs, fables, and parables. Otis included a proverbs test in his original intelligence examination. In this test one has to match proverbs and matter-of-fact statements which might explain them. Kohs (34) used a similar test in his "Ethical Discrimination Test." Kohs, however, used the multiple-choice form, as in the following example:

- People who live in glass houses must not throw stones, means
- () Do not put all your eggs in one basket
 - () Those who have faults should not criticize others
 - () An hour may destroy what it has taken years to build.

In the "Binet-Simon Intelligence Scale" there is a fables test. In the "Stanford Revision" the following fables are used: "Her-

cules and the Wagoner," "The Milkmaid and Her Plans," "The Fox and the Crow," "The Farmer and the Stork," "The Miller, His Son, and the Donkey." After the fable is read to a child, he is asked, "What lesson does that teach us?" Lowe and Shimberg (35) tried out this test as a moral judgments test.

The "Drew Tests in Religious Education" by C. F. Chassell-Cooper (12) represent one effort in this direction. The first of these is a test of the ability to interpret ten parables, viz., "The Lost Sheep," "The Lost Coin," "The Prodigal Son," "The Good Samaritan," "The Sower," "The Ten Virgins," "The Rich Fool," "The Rich Man and Lazarus," "The Unmerciful Servant," "The Pounds." Multiple-choice questions are asked about each parable. There are also self-rating and teacher-rating charts as part of these tests.

Moral Judgment Tests

The "moral judgment" or "ethical perception" test has had a long history. Starting, perhaps it may be said, with the comprehensive work of Sharp, this type of test has suggested the efforts of various workers, including Fernald, Haines, Healy, Watson, and others. Formerly the essay type of examination question was asked, but more recently the whole movement has had a renaissance, using the newer objective testing techniques. Of the older tests, that by Fernald (20) has had the greatest use and publicity. His method has been experimented with considerably in Germany. The illustration given here is taken from Sharp's earlier work (41). It is interesting to note that Sharp used his tests to help throw light on the issue as to whether we solve a moral situation by means of general principles or by noting the effects of a course of action, an issue which is still unsolved and acute to-day. Sharp used such questions as the following (41, p. 202):

(a) In a small Western village a switchman was just about to turn the switch for an approaching express train when he saw his little son, his only child, playing upon the track. The choice had to be made between the life of the babe and the lives of the passengers. What ought he to have done? (b) In the case just cited the man was on duty. What should be the decision under the following circumstances?

A drunken switchman has left the switch open. A man who

Tests of Conduct Knowledge and Judgment 269

lives near the tracks notices the green switch on his way home from work and is just about to turn it to save the train, when he sees his only child upon the track just in front of the engine. The alternative is as in (a).

With the advent of the objective testing movement, tests of moral knowledge or judgment or discrimination employ questions that are simpler and more objective. Since the test items are now shorter and simpler, it is impossible to set up such a heart-rending moral dilemma as is illustrated in Sharp's question. The moral situations in the new-type tests are simpler. Since many more items can be given in the same testing time than were employed for the old examinations, it is believed that ability to make moral discriminations is better tested by the wider sampling of problems, individually not so nerve-racking. A similar change in point of view has taken place in the testing of school subjects. To-day, instead of an examination consisting of difficult problems in algebra, several shorter and easier questions are used in place of each difficult one, with even better result.

Still another difference in the newer testing techniques is the emphasis on situations that actually enter into a child's experience. Sharp's situations were possible but imaginary. The following questions devised by Hartshorne and May show the new point of view (47, p. 17).

1. If another pupil wants to copy your work and hand it in,
 - a. Let him do it and say nothing about it.
 - b. Let him do it and tell the teacher.
 - c. Don't let him do it and say nothing.
 - d. Don't let him do it and tell the teacher he wanted to.
 - e. Don't let him do it and tell him you disapprove of cheating.
5. If you see broken glass in the street,
 - a. Pick it up.
 - b. Do nothing about it.
 - c. Tell the policeman about it.
 - d. Try to find the one who did it.

Lastly the newer type of test uses recognition rather than recall. Instead of asking the pupil to put in writing what he would do, it gives several possible answers out of which the pupil chooses the one he thinks best. This has the advantages of objectivity in scoring and amenability to statistical treatment. The

270 Diagnosing Personality and Conduct

older studies using essay-type tests were handicapped in that the answers were so varied as to preclude classification and statistical treatment. Usually some rough grouping was effected, but the results were subjective and not very satisfactory. What the newer tests lack in permitting free expression they more than compensate for by greater objectivity and reliability.

By far the best tests of this type have been made by Hartshorne and May in their work on the Character Education Inquiry. These tests will be described in some detail and will illustrate the various types used by other authors. The form of the test may vary however.

These authors divided their tests into three groups:

A. *Word Tests*

B. *Sentence Tests*

C. *Good Manners Test* (already described on page 264)

The word tests are four in number, namely, an Opposites Test in the multiple-choice form; a Similarities or Cross-Out Test; a Word Consequences Test, in which the subject indicates (a) *all* likely consequences that might follow from the action represented by the word in capitals; (b) the *most likely* consequence; (c) the *best* consequence; and (d) the *worst* consequence; and the "Schwesinger Ethical-Social Vocabulary Test."

The Word Consequences Test is important for measuring the ability to judge the outcomes of certain situations or actions. Franzen in his series of "Health Education Tests" has used the matching device to achieve the same result, as shown in the following sample: *

- | | |
|------------------|--|
| () Wet feet | 1. Keep from breeding. |
| () Bad cold | 2. Should not touch other people's food. |
| () Bedroom | 3. Blow the nose gently, not hard. |
| () Garbage pail | 4. Keep covered. |
| () Flies | 5. Scald with boiling water. |
| () Sore throat | 6. Should be very clean. |
| () Babies' milk | 7. Should not be too warm. |
| () Sick people | |
| () Whiskey | |
| () Dirty dishes | |

* Reprinted by permission of the American Child Health Association, New York City.

Tests of Conduct Knowledge and Judgment 271

The sentence tests of Group B in the Hartshorne-May series particularly interest us here.

Cause and effect. This is a true-false test of 100 items of which the following are samples (29, p. 6):

Some of the statements made below are true and some are false. Read each statement carefully and underline the word **TRUE** if it seems to you to be true. Underline the word **FALSE** if it seems to you to be false.

- | | | |
|---|------|-------|
| 1. Good marks are chiefly a matter of luck. | True | False |
| 2. Ministers' sons and deacons' daughters usually go wrong. | True | False |
| 3. If one eats stolen apples he will have a stomach-ache. | True | False |

This test is an endeavor to determine how correctly people can reason about cause-effect relations in everyday life. The investigators used the judgments of a graduate class in the psychology of character who answered the items to determine the key. A 75 per cent agreement of this group was necessary before the item was included in the test, with the exception that in one or two cases either ignorance or conventional opinion seemed to determine the answers. In these cases the majority vote of the class determined the key, but in one or two cases the class decision was even revised.

Duties. This is a yes-no test. In it the pupil is asked to state whether certain acts are his duty, are not his duty, or are sometimes his duty and sometimes not (29, p. 7):

- | | | | |
|--|-----|---|----|
| 1. To help a slow or dull child with his lessons | Yes | ? | No |
| 2. To read the newspapers every day | Yes | ? | No |
| 3. To call your teacher's attention to the fact if you received a higher grade than you deserved | Yes | ? | No |

In this test again the question of scoring was difficult. The judgment of the graduate class was even more split than in the previous case and often went against the judgment of the investigators. The majority opinion of the children might also serve as a basis for a key, but in that case it would differ from the key prepared by adults. The preparation of a key for this type of test strongly emphasizes the degree to which these are matters of convention and code.

272 Diagnosing Personality and Conduct

Franzen (21) uses two other devices to test "duties." One he calls a "Five Rules Test." Ten acts of more or less importance are given in a column. The directions are: "In this column mark with an (X) FIVE rules you would like to follow for one month. In this column mark FIVE rules and no more." Since Franzen has interpolated some duties that are not specially related to health, his test is also a measure of something like interest in health or the relative importance with which health is held.

Franzen's other test which belongs in this group is called "Time Test." Twenty-two health acts are listed. The directions state:

- (A) Mark with an (A) the things in this list you should ALWAYS DO ONCE A DAY OR MORE.
- (B) Mark with a (B) the things in this list you should do ONCE A WEEK, ONCE A MONTH, OR MORE OFTEN.
- (C) Mark with a (C) those things in this list you should do ONCE OR TWICE A YEAR.
- () Do NOT mark those things that you do NOT NEED TO DO.*

Comprehensions. This is a multiple-choice test of the "what should one do" type.

Examples (29, p. 8):

- 1. If some one asks to borrow your pencil,
 - (a) Tell him it's broken
 - (b) Tell him that you just lost it
 - (c) Tell him that you don't want to loan it
 - (d) Let him take it.
- 2. If some one steals your lunch,
 - (a) Steal another lunch to even it up
 - (b) Report it to the teacher
 - (c) Cry about it
 - (d) Say nothing about it.

Provocations. This is a more complicated test which attempts to determine the extent to which judgment has the ascendancy over wishes, prejudices, emotions, conventions, and the like.

Example (29, p. 8):

Here are some little stories of what some children did. You are to decide whether they did right or wrong. If what they did

* Reprinted by permission of the American Child Health Association, New York City.

Tests of Conduct Knowledge and Judgment 273

was not quite right, perhaps it was at least excusable in view of the circumstances. Look at the sample first.

Sample: Jane's family was too poor to buy fruit for her sick brother. So every now and then Jane took an apple or an orange from a fruit stand and brought it home to him.

Now if you think she was absolutely wrong in taking the fruit, put a circle around the *Wr*, like this

R Ex **(Wr)**

But if she did exactly right, encircle the *R*, like this

(R) Ex Wr

If you think she was wrong, but excusable in view of her desire to bring it to her sick brother, encircle the *Ex* like this

R **(Ex)** Wr

Begin here and do the rest in the same way.

1. Helen noticed that nearly every one in the class was cheating on a test so she cheated too.
2. Harry was a Christian boy. One day a Jewish boy called Harry "a dirty Christian." Harry knocked him down.
3. Charles did not want to play marbles for keeps but the boys called him a "sissy" so he went ahead and played for keeps anyway.

R Ex Wr

R Ex Wr

R Ex Wr

In this test it was found that the responses of the graduate class tended to be so highly conventional—much like what the sixth-grade children answered—as to be worthless. For a scoring key the examiners used answers which agreed with "a standard that would conform to the great historical moral ideals."

Foresights. This test consists of a number of descriptions of situations. The subject is merely requested to write down as many things as he can think of, good and bad, that might happen from the events recorded (29, p. 10):

1. Whenever any one picked on John he would go tell his teacher.
2. John accidentally broke a street lamp with a snow-ball.

In a later edition this test is given in the multiple-choice form in which a number of possible consequences are listed and the pupil is asked to rate each one according as "this is likely to happen"; "this might happen but not likely"; and "this would not happen."

274 Diagnosing Personality and Conduct

<i>This is likely to happen</i>	<i>This might happen but not likely</i>	<i>This would not happen</i>	<i>John accidentally broke a street lamp with a snow-ball</i>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1. John was arrested and sentenced to six months in jail.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2. John said nothing about it, and people thought another boy had done it.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	3. The emergency wagon had to come and repair it.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	4. He thought it was such fun that he smashed a lot more lamps.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	5. There was an accident there because it was dark.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	6. Some people were cross about it and John's father got into trouble.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	7. The glass went on the street and a child cut his hands on it.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	8. The city had to pay for the lamp.

Recognitions. This is a test of ability to classify acts under certain headings. The directions state (29, p. 10):

After each statement are five letters, C L S X J. If the deed is a case of Cheating, draw a circle around the C; if it is Lying, around the L; if it is Stealing, around the S. If it is something wrong, but not either cheating, lying, or stealing, put a circle around the X. If it is not wrong at all, put a circle around the J. If the thing is both cheating and lying or stealing and lying, or all three, encircle all the letters you need to in order to express your opinion.

- | | |
|--|-----------|
| 1. Bullying younger children | C L S X J |
| 2. Using street-car transfers that are out of date | C L S X J |
| 3. Riding on the back of a truck without the driver's knowing it | C L S X J |

Principles. Whatever one's belief or theory may be as to the nature of rational action, it is certain that on some occasions we

Tests of Conduct Knowledge and Judgment 275

act not on the basis of foreseen consequences, but on the basis of principles. The next test in the C. E. I. series is a true-false test of knowledge of principles (29, p. 11):

- | | | |
|---|------|-------|
| 1. To master oneself is a greater thing than to win a battle. | True | False |
| 2. Clean speech is a sign of being a "goody-goody." | True | False |
| 3. Obedience is of greater importance than honor. | True | False |

Applications. Next comes a test of the ability to apply these principles. A situation is described which ends in a dilemma in conduct, requiring the making of a choice. Five principles are then given, two of which apply in the situation. These two are so chosen that one suggests the first alternative as the right one, and the other the second. If a pupil has his decision already formed, he will not be so ready to recognize a principle which applies in favor of the opposite side of the issue.

Example (29, p. 12):

1. Mary saw Helen cheating on an examination. She had to decide whether she would

- () (a) Report it to the teacher.
 () (b) Not report it to the teacher.

Here are the five rules, of which two apply to this problem. Check two and only two in the spaces at the left of the numbers.

- () (1) Treat others as you would like to have them treat you.
 () (2) Be true to what is for the good of all, even when your own interests or those of your friends are involved.
 () (3) When you have wronged some one, ask to be forgiven.
 () (4) Be cheerful and uncomplaining when disappointed or hurt or in trouble.
 () (5) Do not think of yourself as more important than you are.

After checking the two rules that apply to Mary, put a check before either (a) or (b), according as you think it would have been right for her to tell or not to tell.

In later revisions a further test is given which carries this last test a step further. Instead of asking whether or not the

276 Diagnosing Personality and Conduct

principles apply, the pupil is asked to state which principles are most important. For each incident the principles for and against are listed separately. The test is headed "What Things Would Be Important If They Were to Happen."

In the test the pupil is to put a circle around all of the principles he thinks are important in the situation described. Then he is to put an X before the two most important principles, and finally a second X before the most important of all.

Example:

1. John accidentally broke a street lamp with a snow-ball.

First Ballot

- a. The glass fell and hurt a man who was passing by.
- b. His father scolded John for it.
- c. John blamed another boy for it.
- d. It showed him that he had to be more careful when throwing snow-balls.
- e. The city had to pay for the lamp.

Second Ballot

- f. The emergency wagon had to come and repair it.
- g. A policeman saw John do it and ran after him.
- h. The glass went on the street and a child cut his hands on it.
- i. John said nothing and people thought another boy had done it.
- j. John paid for it.

Altogether these tests represent a rather complete analysis of the different phases of the thought process as applied to conduct. There are many theories of the relation of thought to conduct. On the one hand some believe that much or most of conduct operates without the aid, guidance, or intervention of thought. Others believe that thought acts through the application of general principles. Still others believe that thought penetrates deeper, first sizing up the situation, afterwards considering what the outcomes may be, and then deciding in the light of these outcomes.

The tests called "Duties," "Comprehensions," and "Provocations" get at the knowledge of what one ought to do or judgment as to whether an act is right, wrong, or excusable.

Tests of Conduct Knowledge and Judgment 277

To test out the control of conduct on the basis of general principles several tests are available. The "Recognitions" test determines ability to classify acts under certain headings. This is important if one is to be able to apply the principles to specific situations. Then there is a test of knowledge of "Principles" and a test of the ability to "Apply" these principles in the concrete situation. Finally there is a test of ability to "Weigh" and "Choose" these principles according to their importance.

To test the ability to guide conduct by noting its outcomes or results three tests are available. Two of these, the "Word Consequences" and the "Foresights" tests, measure the ability to estimate the probable outcomes of acts or situations. The "Cause and Effect" test is more a test of looking back from a given event to the true and proper cause.

One other test seems necessary in this latter process—a test of the ability to weigh consequences and make choices in the light of the decision. Such a test has been prepared by Chassell and Chassell (11) in the "Test of Ability to Weigh Foreseen Consequences." The test consists of a series of stories. After each story is given a list of ten or so *possible consequences*. The directions state to "mark with a plus sign (+) all the consequences that seem to you desirable, and with a minus sign (—) all those that seem to you undesirable." After this is done the pupil is asked to make a decision in the light of the consequences and their importance.

Underline YES if the consequences marked plus seem to you the more important, and NO if those marked minus seem to you the more important.

YES

NO

This test is important, and strangely enough it supplements the C. E. I.'s excellent battery of tests in an important way. The test as it stands may be criticized because the stories describing the incident to be judged are so lengthy. Reading ability, both speed and comprehension, certainly play a large part in determining outcomes on the test. In the C. E. I. tests, all of the incidents are described in a very brief way.

Another very interesting and suggestive test is the "Story Test" devised by Franzen as one of the "Health Education Test" series.

278 Diagnosing Personality and Conduct

It is hard to classify this test as one of knowledge, interest, attitude, judgment, or conduct. A story is read and the pupil is to note with a red pencil all the items relating to health, good or bad. Since the stories themselves are interesting, providing a good deal to distract the attention from the health items, a child must overcome a certain amount of resistance to turn his attention from the thread of the story to the matters of health. Exactly what the test measures must be learned from the correlations.

Example: *

DIRECTIONS

Underline everything that is good for health.

Cross out everything that is bad for health.

Use a red pencil.

Two things are already marked correctly in this story.

Jean Learns How to Play a New Game

Jean ran down the long hall through the school toward the playground. As she passed the drinking fountain she stopped for a long cool drink. She ~~pressed her lips~~ ~~close~~ to the pipe so as not to waste a drop of the sparkling water. The girls in Jean's gymnasium class were to learn to play basket-ball that afternoon, and they were very much excited about it. They played with the ball, trying to throw it into the basket, until their time in the gymnasium was over. When they were leaving the teacher gave each one a book of rules to look over at home before they played again.

Several of Jean's classmates lived near her and on their way home from school they planned to play basket-ball all afternoon. One of the girls borrowed a basket-ball from her brother and Jean's brother fastened a barrel hoop to a tall fence for them to use as a basket. The girls had a wonderful time and yelled and shouted until they were hoarse and very thirsty.

They all went into Jean's house to have a drink and counted out to see who should be first to use the pretty green glass that stood near the faucet. Jean's mother gave each one three cookies and the girls sat on the porch to eat them and plan for other afternoons practising with the ball. One of the girls thought her father might give her a basket-ball for her birthday. Jean's mother had told them when she gave them the cookies that it was five o'clock so they could only stay a few minutes longer.

After supper Jean took her book of basket-ball rules and sat

* Reprinted by permission of the American Child Health Association, New York City.

down to read in front of a lamp so that her book was in a bright light. She studied them carefully and by eight o'clock she was sure she would be able to play the game the next time they went to the gymnasium.

Jean was very tired and was glad she didn't have any more studying to do so that she could go right to bed.

Offense Rating

In this group of tests of knowledge and judgment the ranking or rating of offenses should be included. In the tests previously described, a single judgment was required on a variety of items, each one taken separately. In the present technique a careful ranking of a small group of carefully chosen items is asked for instead. The method of ranking or paired comparisons has had a long psychological history. Fernald was the first to suggest that this method be used in ranking offenses, so far as I am aware.

Brogan (3, 4; 5), who has studied the method somewhat extensively, asked students in the University of Texas to list the "ten worst practices" among students in the university. Sixteen practices were found as occurring most frequently. They are (in alphabetical order): Cheating, Dancing, Drinking, Extravagance, Gambling, Gossip, Idleness, Lying, Sabbath-Breaking, Selfishness, Sex Irregularity, Smoking, Snobbishness, Stealing, Swearing, Vulgar Talk. Brogan has had these lists ranked many times by various groups according to the "worseness" of the practice, and according to the *frequency* with which it occurs.

The following tests by Cushing and Ruch (19), similar to a test of the same nature by Raubenheimer (37), who in turn drew his material from the "Clark Rating Scale of Offenses," shows the trend toward using described situations with which younger boys would be more familiar.

Directions

Below are ten offenses committed by boys in a certain reform school. Read them through carefully and find the worst offense. Mark it 1. Read them through again and find the next worst offense. Mark it 2. Mark the next worst 3, and so on down to the least serious, which will be 10. Every rank from 1 to 10 must be used. No tie ranks are permitted.

280 Diagnosing Personality and Conduct

Begin here:

	<i>Rank</i>
(a) Associated with hoboes—slept out in caves, tents, and boxes.	_____
(b) Burned the public school in Los Angeles, which he attended.	_____
(c) When his father started to take him to a detention home for misconduct, he asked to return to the house to say good-bye to his mother. There he took a 22-rifle and shot his father, killing him instantly.	_____
(d) Ran away from home and obtained employment, securing his room and board with another family.	_____
(e) Forged a check for \$10.	_____
(f) Played hookey to attend a circus.	_____
(g) Together with another boy was accused of murder. They had been drinking, were given more liquor by two adults, and in a fight which ensued one of the men was killed.	_____
(h) Stole scrap iron from railway cars on a siding and sold it to a junk man.	_____
(i) Was sent to a parental school because he struck his teacher. Was accused of drawing a knife on a boy whose bicycle he had stolen, when the boy claimed it. Arrested for carrying a blackjack and knife.	_____
(j) Entered the house of the next-door neighbor and took \$2.50.	_____

Cady uses rating instead of ranking. The relative advantages of rating and ranking are discussed on page 76. His plan is to have each quality or offense or activity rated on a four-place scale according to the amount of blame one would assign to it. His directions and a sample of the test follow (10, p. 135):

Here is a list of eight words which describe people who are grouchy, extravagant, nervous, forgetful, etc. Divide such people into four classes by placing a cross opposite each word according to whether you think them:

- (1) very greatly to blame.
- (2) a good deal to blame.
- (3) a little to blame.
- (4) not at all to blame.

Look at each word. Think how you feel about such people and then place them in one of the four classes.

Tests of Conduct Knowledge and Judgment 281

<i>people who are:</i>	(1) <i>very greatly to blame</i>	(2) <i>a good deal to blame</i>	(3) <i>a little to blame</i>	(4) <i>not at all to blame</i>
<i>grouchy</i>				
<i>extravagant</i>				
<i>nervous</i>				
<i>forgetful</i>				
<i>dishonest</i>				
<i>careless</i>				
<i>impulsive</i>				
<i>unambitious</i>				

Still another technique is the cross-out test of Pressey. In this test a list of words is given and the subject is instructed to cross out all the words which stand for things he thinks are bad.

Illustration:

Directions: Read through the twenty-five lists of words given just below and cross out EVERYTHING THAT YOU THINK IS WRONG—everything that you think a person is to be blamed for. You may cross out as many or as few words as you like; in some lists you may not wish to cross out any words. Just be sure that you cross out everything you think is wrong.

- | | | | | |
|-------------|----------|-----------|----------|-------------|
| 1. begging | smoking | flirting | spitting | giggling |
| 2. fear | anger | suspicion | laziness | contempt |
| 3. dullness | weakness | ignorance | meekness | stinginess. |

The matter of scoring a test which is in the form of ranking has always given some trouble. Raubenheimer used as a scoring method for his test of offense rating "the sum of the square of the deviation in ranking for each item from the standard ranking developed by Clark in his 'Offense Rating Scale.'" The advantage in squaring the differences is that it increases the reliability of the scores.

Naturally the lower the score in this test, the closer the agreement with the standard.

Cady brings out the point that in ratings standards vary considerably, so that when one tries to assign a score to an individual on the basis of his deviations from some norm or stand-

TABLE 49

SCORING JUDGMENTS EXPRESSED AS RANKS

Item	Standard rank	Rank given by a pupil tak- ing the test	Deviation	Deviation squared
a	3	4	1	1
b	8	7	1	1
c	10	8	2	4
d	2	5	3	9
e	6	6	0	
f	1	3	2	4
g	9	9	0	
h	4	1	3	9
i	7	10	3	9
j	5	2	3	9
				<hr/>
Sum				46

ard, his variations may be large not because he judges carelessly or inaccurately but because he holds radically higher or lower standards. What one wants is not the average amount of deviation of the ratings from some norm, but the average amount of the deviations from their own mean. Kelley* gives for the computation of the average deviation a short formula which is convenient to use in this problem. The formula is

$$\text{Av. Dev.} = \frac{2}{N} \left(FM - \sum_{i=1}^F X \right)$$

where N = number of cases (things rated)

F = number of cases lying below the mean

M = mean

$\sum_{i=1}^F X$ = sum of the ratings below the mean

* Kelley, T. L., *Statistical Method* (The Macmillan Company, 1923).

Tests of Conduct Knowledge and Judgment 283

<i>True values as given by compe- tent judges</i>	<i>Ratings assigned in test</i>	<i>Deviations from the true values</i>
6.5	6.5	0
2.6	3.0	+ .4
7.9	7.0	- .9
3.2	4.5	+1.3
0.8	2.0	+1.2
4.5	4.0	- .5
1.25	1.7	+ .45
1.1	1.5	+ .4
8.0	6.5	-1.5
2.2	2.8	+ .6
Sum of absolute values of de- viations		$\frac{7.25}{10} = .725$ av. error
Sum of algebraic values of deviations		$\frac{1.45}{10} = +.145$ { av. am't by which rat- ings were too high

Calling this value, $+.145$, the mean, there are 4 scores (0, $-.9$, $-.5$ and -1.5) lying below this mean.

$$\begin{aligned} N &= 10 \\ F &= 4 \\ M &= +.145 \end{aligned}$$

$$\sum_{i=1}^F = 0 - .9 - .5 - 1.5 = -2.9$$

$$\begin{aligned} \text{Av. Dev.} &= \frac{2}{N} \left(FM - \sum_{i=1}^F X \right) \\ &= \frac{2}{10} \left[4 \times .145 - (-2.9) \right] \end{aligned}$$

$$= \frac{2}{10} (.580 + 2.9)$$

$$= \frac{3.480}{5}$$

$$\text{Av. Dev.} = .696$$

If the work is carefully planned, this computation may be readily run off, and the "score" of each individual may thereby be obtained with his tendency to overrate or underrate eliminated.

Reliability

Reliabilities in this testing have not been very frequently reported. Chassell and Chassell (11) give the reliability of their Test of Ability to Weigh Foreseen Consequences as .91 when obtained for fifty-nine junior high school pupils. Raubheimer (37) gives a reliability coefficient of .78 for his offense rating test in a thirteen-year-old group. The best reliabilities are those reported by Watson (47) and the C. E. I. (29).

Watson's reliability coefficients are given in the following table:

TABLE 50
RELIABILITY COEFFICIENTS OF TEST OF CONDUCT INFORMATION AND OPINION
(from Watson, 47, p. 36)

<i>Subject and type of test</i>	<i>No. of elements</i>	<i>Reliability coefficient (half against half)</i>	<i>Self-correlation for 100 similar elements</i>
True-false statements on religious ideas.	20	.45	.80
"Duties" test—multiple-choice	20	.48	.82
Yes, doubtful, no test on practices.....	133	.90	.88
True-false test on leadership, obedience, patriotism, etc.	10	.90	.99
Ranking ways of acting in groups of three	15	.85	.97
Choose best and worst of five alternatives dealing with problems of boys..	8	.70	.96
Check items in list which tend to pervert one	17	.79	.95
Choose best and worst among five alternative ways of acting	8	.35	.87
Checking + ? or — methods of showing Christian spirit in home duties....	11	.21	.70
Yes ? No test on best ways of getting on at school	10	.08	.46

Hartshorne and May give the following reliabilities for their tests:

TABLE 51
RELIABILITY COEFFICIENTS—MORAL KNOWLEDGE TESTS

	<i>Coefficient of reliability</i>
Opposites828
Cause-effect778
Duties832
Comprehensions805
Provocations733
Recognitions798
Principles688
Applications of principles810

Schwesinger gives the exceedingly high reliability of .98 for the Ethical-Social Vocabulary Test (40).

In general these are satisfactory reliabilities and are only slightly less than one expects from tests of the same type in school subjects. The reliabilities are fully as high as are obtained elsewhere with questions that are matters of fact or general agreement as in the vocabulary test; but they are a bit lower where codes and standards vary, as is also true of tests of English usage. Indeed one might well wonder that the tests show as high consistency as they do in view of the fact that ethical standards depend so largely on a variable point of view.

Validity

The question of exactly what significance the various tests of knowledge, judgment, and discrimination in the field of conduct have is of great importance. Undeniably these tests have been constructed and used in the belief that they have some relationship to actual conduct. There is strong popular conviction that knowledge and reasoning exercise potent control over conduct. If this is true, these tests are of supreme importance, and even though it is true that language tests do not have direct relationship to specific modes of behavior, they may represent whatever integrating force there is for behavior as a whole. In other words, even though language tests may not correlate with specific modes of conduct, they may be very important as indices of general character. What are the facts?

286 Diagnosing Personality and Conduct

Intercorrelations of moral knowledge tests. By far the most significant work in the validation of conduct knowledge tests is that of Hartshorne and May. They give intercorrelations for their tests as follows:

TABLE 52
INTERCORRELATIONS OF C. E. I. MORAL KNOWLEDGE TESTS
(from Hartshorne and May, 29, p. 18)

	2	3	12	4	5	6	7	9	10	11	X
1. Opposites...	x	x	.748	.750	x	-.362	.197	x	.572	x	...
2. Similarities..	..	x	.612	x	x	.236	x	.472	x	x	...
3. Word consequences...665	x	x	.137	x	-.440	x	x	...
12. Vocabulary..458	.383	.381	.276	.389	.330	x	.720
4. Causes.....350	.000	.237	.500	.555	.326	.174
5. Duties.....228	.331	.600	.575	.030	.302
6. Comprehensions.....000	.400	.430	.363	.443
7. Provocations172	.248	.463	.396
9. Recognitions276	.258	.478
10. Principles...164	.438
11. Applications274
X. Good manners

One should first note that all the correlations are substantially positive, indicating that there may be some common factor running through all of the tests. The logical grouping of tests which was made on page 277 finds no counterpart in the correlations. Tests 5, 6, and 7, which test more or less directly knowledge of desirable conduct, correlate .228, .331, and .000 respectively. Tests of knowledge of moral principles and their applications (9, 10, and 11) also have low correlations (.276, .258, and .164). On the other hand the vocabulary tests have substantial intercorrelations. The correlations perhaps are best explained by the mechanical nature of the tests and the psychological kinship of the functions rather than by any logical relationship which might be imagined. The reliability of the tests may be one factor determining the size of the intercorrelations. The matter of standards of the answers may be another factor. Tests in which there is common agreement as to the rightness or wrongness of an answer should correlate better than tests in which the standards are more subjective. This may partly explain the higher correlation of the

vocabulary tests. Again, tests of knowledge, even though they may merely mirror current codes of behavior (as the tests of duties and principles, .575) and the tests of judgment and application (such as the tests called provocations and applications, .463) may correlate better than tests in one of these groups with tests in another (principles and applications, .164). If these are the factors which determine the correlations, it is evident that they are not very helpful in deciding upon the possible dynamic relations between the various phases of knowledge and reasoning on conduct. But they do indicate that the different types of activity are loosely strung together, making it possible for a person to possess principles of conduct even though he is not in a position to estimate probable outcomes in action or to apply the principles.

Correlations of moral knowledge tests with a composite. Hartshorne and May also found the correlations of their Moral Knowledge Tests with the weighted composite of the first seven tests. Assuming that the reliability of the composite is .90, these correlations, corrected for attenuation, are as follows:

TABLE 53

CORRELATION OF JUDGMENT AND REASONING TESTS AGAINST A COMPOSITE OF
VOCABULARY AND KNOWLEDGE TESTS

(from Hartshorne and May, 29, p. 21)

Causes521
Duties544
Comprehensions372
Provocations421
Recognitions581
Principles636
Applications418

Correlations of moral knowledge tests with intelligence. Healy and Fernald (30) concluded from their use of a moral judgments test that it was useful for estimating a subject's powers of intellectual comprehension. Chassell and Chassell found their test to correlate with IQ .43 and MA .49. Raubenheimer's test of offense rating correlated with the National Intelligence Test .63. Hartshorne and May find that their moral knowledge tests correlate with intelligence, as determined on intelligence test material furnished by Thorndike, as follows:

TABLE 54

CORRELATION OF MORAL KNOWLEDGE AND INTELLIGENCE

(from Hartshorne and May, 29, p. 20)

Opposites775
Similarities664
Word consequences519
Cause and effect647
Duties402
Comprehensions371
Provocations145
Recognitions498
Principles444
Applications562
Vocabulary882
Composite of seven tests....	.686

These are substantially high and indicate that general intelligence is a large factor in determining a score on a test of moral knowledge. This is further emphasized by noting that when intelligence is partialled out of the correlations between a moral knowledge test and the composite of moral knowledge, the correlation drops down close to zero. "Duties," "provocations," and "principles" seem to test something beside intelligence, however.

Correlations of moral knowledge and age. Hartshorne and May found the following correlations between age and certain tests.

TABLE 55

CORRELATION BETWEEN MORAL KNOWLEDGE TESTS AND AGE

(from Hartshorne and May, 29, p. 22)

Opposites	-.094	Comprehensions416
Similarities	-.068	Provocations	-.097
Word consequences	-.204	Recognitions172
Causes473	Principles026
Duties138	Applications183
Vocabulary	-.091		

From these figures it would seem that, in general, moral knowledge shows little progression with age for children in grades five to eight; much less change, in fact, than is usually found with school subjects. It would seem as though moral knowledge becomes fixed rather early in life or that it is relatively uniform throughout a group such as a class or school. Slavons and Brogan (43) in studying judgments of the seriousness of acts, found that

high school and college students hold approximately the same moral standards, from which finding he reaches the conclusion that these standards become fixed relatively early.

Correlation of moral knowledge and conduct. The testimony here is remarkably unanimous. In 1911 Healy and Bronner (30) concluded that moral judgment has little relation to a subject's moral nature. Haines (24) found that a moral judgment test did not differentiate between normal and delinquent girls. Lowe and Shimberg (35) discovered that the Binet "Fables Test" does not differentiate delinquents from non-delinquents. Schwesinger (39) reported that her test of slang does not separate delinquents from non-delinquents. Raubenheimer (37) found that the test of offense rating differentiated between delinquent and non-delinquent groups to a degree represented by correlations of .31, .26, .24, and —.11. Cushing and Ruch (19) found the biserial r between delinquent and non-delinquent groups and offense rating to be .27. Weber (48) reported a high correlation between the order of Brogan's items rated according to seriousness of offense for delinquent and normal women. Branham (2) believes from his study of criminals that the defective delinquent can tell more readily whether an act is right or wrong than he can tell the enormity of the offense. Chassell and Chassell (11) found a correlation of .17 between their test of ability to judge consequences and ratings on conduct.

Hightower, from his study of the relation of Biblical information to character and conduct, concludes that a knowledge of the Bible has no relationship to any phase of the conduct which he tested. His study indicates that mere knowledge is not of itself sufficient to insure proper character growth.

Hartshorne and May (27), in the most analytical work performed in this field, present correlations found in Table 56 on the following page between tests of moral knowledge and cheating.

These correlations are very low. The conclusion is that *there is a very low positive correlation between moral knowledge and desirable conduct*. When groups socially selected on the basis of conduct are tested with moral knowledge, this relationship is also so low as to be distinguishable only with reliable tests. These conclusions have far-reaching implications. They end once for all speculation as to a general relationship between conduct and knowledge. Hartshorne and May even go further and say that it

TABLE 56

RELATION OF CHEATING TO MORAL KNOWLEDGE
(from Hartshorne and May, 29, p. 25)

<i>Tests</i>	<i>Home cheating</i>	<i>School cheating</i>
Cause and effect	+.031	-.054
Duties	-.178	-.296
Comprehensions	-.018	-.301
Provocations	-.129	-.241
Recognitions	-.091	-.181
Principles	-.088	-.247
Applications	-.066	-.402
Sum of 1-7	-.121	-.385

is not possible to pick out items of moral knowledge which consistently predict or correlate with conduct.

So much for the bare facts of relationship. Various investigators have followed the matter further and have discovered other facts and relationships that help uncover the basis for conduct knowledge and judgment.

Sex differences. The sexes hold the same standards with respect to conduct to a remarkable degree. Brogan (4) found a high degree of agreement in the rating of offenses according to seriousness, except as regards *smoking* and *idleness*. Probably the differences between the sexes in the matter of smoking has broken down to some extent since Brogan did his work; and in the matter of idleness the difference will also grow less as women enter industry, compete with men, and occupy less and less the position of an idle class. But the conclusions of Snyder and Dunlap (42) and of Tanaka (43) are that women are more severe than men in expressing their judgments on conduct issues. The bad they condemn more harshly and the good they praise more highly.

Ethical standards. The matter of standards is very troublesome. In some fields such as health, standards are fairly definite, but for a moral knowledge test there are no answers which may be commonly accepted as standard, as there are for a test of knowledge in history or science. Who is to be the arbiter? Hartshorne and May found that the judgment of a graduate class in character measurement often resembled the responses of a class of children more than it did traditionally great historical moral ideals. If *to test* means merely "to survey," then the tests cannot be scored in the usual sense. Every answer must be taken at its face value

But if we grant that there are standards and that some kinds of conduct are better or more important than others, then the test may be scored. Hartshorne and May wondered whether children may not have their own conduct codes as contrasted with the adult or ideal standard, and whether this child code may not have some positive relation to their conduct. They made a key for all their tests based on children's typical or modal responses. The correlations of four of the tests with deception as measured by performance tests were $+.015$, $+.071$, $-.058$, and $.115$. These are even closer to zero than when the ideal key was used, indicating that the matter of standards is not the factor which eliminates the relation with conduct. In the social-moral area, the best available expert judgment must be used to determine standards.

The personal or abstract answer. That one's abstract judgment of what should be done in a moral situation and his statement of what he would do himself may not agree has always been recognized. On the other hand, since the two responses are both verbal, the relation should be considerably closer than the relation between knowledge and conduct itself. Hartshorne and May found a correlation of $.77$ between two tests of duties, one in which the questions are stated:

It is my duty

To help a slow or dull child with his lessons.. Yes ? No

To read the newspapers every day Yes ? No

and the other in which the questions are stated:

Did you ever help a slow or dull child with his lessons? Yes No

Do you read the newspapers every day? Yes No

Cady raises a similar question when he inquires whether the response will be the same when the pupil is thinking of the situation with respect to a large group and when he is thinking of it with respect to a small group. This leads to the larger question of the relation of the response of an individual to the group of which he is a part and with which he has been associated.

Moral knowledge and the conventional code. An old dispute in ethics raged around the issue whether one determines his course of action by a consideration of the consequences of an act or by the application of general principles. The answer that modern experimentation gives to this question is that neither is a cor-

292 Diagnosing Personality and Conduct

rect description of what happens. Habit is the main control of conduct. But there can be no denying that the question is a real one, for we are constantly making decisions, some of which issue into conduct. Our tests make it clear that actual foresight of consequences is a rare method of solution, and the consensus of opinion of those who have studied tests and their results is that children's knowledge and judgment of conduct reflect the conventional code or codes. Snyder and Dunlap (44), who tested college students on their judgment reactions to a large number of personal experiences, conclude that there is sometimes a conflict between the traditional and the reflective as evidenced by the high variability of response, but that the traditional dominates over the reflective except in cases where the welfare of others is concerned.

Interesting evidence in this connection has been obtained by Hartshorne and May (29), who gave their moral knowledge tests to fathers, mothers, teachers, Sunday-school teachers, and playmates of a certain group and found which groups had a code most like that of the children being tested. The following correlations were reported:

TABLE 57
CORRELATIONS BETWEEN MORAL KNOWLEDGE OF CHILDREN AND OF
OTHER GROUPS
(from Hartshorne and May, 29, p. 43)

Children and father40
Children and mother49
Children and both parents545
Children and teacher028
Children and Sunday-school teacher.....	.002
Children and club leaders137
Children and playmates353

These facts would indicate that children obtain their codes largely at home and from their playmates. It might be said that the influence is proportional to the amount of time spent together and the amount of sharing of experiences. These same investigators found that the codes of children vary according to the group in which they are tested.

Moral knowledge and group conduct. Hartshorne and May have further demonstrated that the relationship between knowledge and conduct is strikingly higher when groups are considered as a whole rather than as individuals.

TABLE 58

CORRELATIONS BETWEEN MORAL KNOWLEDGE AND CONDUCT FOR INDIVIDUALS
AND GROUPS—AVERAGE OF EIGHT TESTS

(from Hartshorne and May, 29, p. 44)

	<i>Individual</i>	<i>Group</i>
Behavior C (undesirable).....	-.25	-.44
Behavior A (undesirable).....	-.13	-.15
Behavior H (desirable).....	+.23	+.53

These facts help to explain the origin of our knowledge of conduct and our conduct codes. Hartshorne and May find in them indications that standards of conduct operate through group activity as opposed to individual activity.

"Since the group r 's are larger than the individual r 's, they cannot be accounted for by a causal relation between moral knowledge and conduct, since this relation could operate only through the minds of the individuals concerned. Hence the superiority of the group r 's must be due to the reaction of individuals to some influence which tends toward higher code and more social conduct (and *vice versa*) without these being integrated in the minds of the individuals. Such a common influence might be exerted by the group as a whole through a growing tradition or by the teacher or by school system, or by all three. No matter how much it affects either conduct or code for the better, if the correlations indicate the absence of individual integration, this improvement can hardly be regarded as growth in character." (29, p. 68.)

Acceptance of this last statement would seem to depend on how character is defined. Its authors have not been uninfluenced by a wish as to what character ought to be, as contrasted with the attitude of taking things as they are found.

Miscellaneous Factors Associated with Moral Knowledge

Snyder and Dunlap (44) in their study of the moral judgments of college students uncovered other factors which determine moral judgment. The magnitude of the act (meaning the seriousness of the consequences, the number of people involved, the amount of damage done) is one factor on which judgment is based. Brogan discovered that there is a negative correlation between the frequency of an act and judgment as to its seriousness. Acts that are frequent will not be judged to be serious.

294 Diagnosing Personality and Conduct

Motive back of an act is also a factor used in judging the seriousness of an act. Usually the motive must be guessed at from the description of the act printed on a test blank. But the motive is equally hidden in actual life situations. It is the writer's opinion that the educational outcome is saner and leads to better integration when greater attention is paid to the social outcomes of behavior than to the motives leading to the behavior.

Other factors which influence judgment are the condition of the person affected by the act, and the degree of personal sacrifice involved in performing the act. Social virtues are more highly esteemed than those involving mere individual well-being. Tanaka (45) found that Japanese children agreed tolerably well in evaluating issues concerning state matters and foreign countries, but that there was considerable difference of opinion in judging personal standards of conduct.

Conclusion

It is possible to measure knowledge and judgment with reference to conduct through the application of several useful tests which have been constructed for measuring health, Biblical knowledge, ethical knowledge, etc. These tests have very satisfactory reliability, comparing favorably with tests in the school subjects, similarly constructed. They correlate somewhat with each other and substantially with intelligence in general. The correlations with conduct are very low, so that with the less perfect tests they seem to fail to differentiate between normal and delinquent individuals. Conclusions from the research work done with these tests seem to indicate that answers reflect the code of the group in which the individual happens to be rather than any reasoned solution to the problem situation presented for judgment. These codes seem to be group affairs, and there is a distinct correlation between conduct and knowledge when groups as a whole are considered. The low correlations of knowledge and conduct for individuals indicate how distinct the two forms of activity are. On the other hand, when these correlations are compared with the correlations between different forms of conduct, there is ground for the suggestion that perhaps knowledge and judgment of conduct constitute after all the one force, however ineffective, that works toward integrating conduct.

REFERENCES

1. BLOMFIELD, F. M., "Church Schools Tested," *Religious Education*, 22: 471-476 (1927).
2. BRANHAM, V. C., "The Classification and Treatment of the Defective Delinquent," *Journal of Criminal Law and Criminology*, 17: 183-217 (1926).
3. BROGAN, A. P., "Group Estimates of Frequency of Misconduct," *International Journal of Ethics*, 34: 254-271 (1924).
4. ———, "Moral Valuations about Men and Women," *International Journal of Ethics*, 35: 105-124 (1925).
5. ———, "A Study in Statistical Ethics," *International Journal of Ethics*, 33: 119-134 (1923).
6. BRONNER, A. F., *Comparative Study of the Intelligence of Delinquent Girls*, Teachers College Contributions to Education, No. 68 (1914).
7. BROTEMARKLE, R. A., "College Student Personnel Problems, I. Individual Mental Testing at the College-Adult Level," *Journal of Applied Psychology*, 11: 415-436 (1927).
8. ———, "A Comparison Test for Investigating the Ideational Content of the Moral Concepts," *Journal of Applied Psychology*, 6: 235-242 (Sept., 1922).
9. BURT, C., *The Young Delinquent* (D. Appleton and Company, 1925).
10. CADY, V. M., *The Estimation of Juvenile Incurability*, Journal of Delinquency Monographs, No. 2 (California Bureau of Juvenile Research, Whittier, California, 1923), 140 pp.
11. CHASSELL, C. F., CHASSELL, E. B., and CHASSELL, L. M., "A Test of Ability to Weigh Foreseen Consequences," *Teachers College Record*, 25: 39-50 (Jan., 1924).
12. CHASSELL, C. F., "Some New Tests in Religious Education," *Religious Education*, 16: 319-336 (1921).
13. CHASSELL, C. F., and CHASSELL, E. B., "A Test and Teaching Device in Citizenship for Use with Junior High School Pupils," *Educational Administration and Supervision*, 10: 7-29 (1924).
14. CHASSELL, C. F., and CHASSELL, L. M., "A Test of Religious Ideas Involving the Ranking of Selected Answers," *Religious Education*, 17: 45-49 (1922).
15. ——— and ———, "Multiple Choice Test of Religious Ideas," *Indiana Survey of Religious Education* (George H. Doran Company, 1925; now published by Institute of Social and Religious Research), Vol. II, ch. 17.
16. CLARK, W. W., "The Measurement of Social Attitudes," *Journal of Applied Sociology*, 8: 345-354 (1923-1924).
17. COLLINS, M., "Character and Temperament Tests," *British Journal of Psychology* (Gen. Sect.), 16, pt. 2: 89-99 (1925).

296 Diagnosing Personality and Conduct

18. COLVIN, S. S., "Principles Underlying the Construction and Use of Intelligence Tests," in *Intelligence Tests and Their Use, Twenty-first Yearbook of the National Society for the Study of Education* (1922), 40-42 pp.
19. CUSHING, H. M., and RUCH, G. M., "An Investigation of Character Traits in Delinquent Girls," *Journal of Applied Psychology*, 11: 1-7 (Feb., 1927).
20. FERNALD, G., "The Defective Differentiating Class: Differentiating Tests," *American Journal of Insanity*, 68: 523-594 (April, 1912).
21. FRANZEN, R., *Health Education Tests* (American Child Health Association, 1929).
22. GATES, A. I., and STRANG, R., "A Test in Health Knowledge," *Teachers College Record*, 26: 867-880 (June, 1925).
23. GILES, J. T., "Sunday School Examination A," in *The Indiana Survey of Religious Education* (George H. Doran Company, 1925; now published by Institute of Social and Religious Research), Vol. II, ch. 14.
24. HAINES, T. H., "Diagnostic Value of Some Performance Tests," *Psychological Review*, 22: 299-305 (1915).
25. HANSON, W. L., "Boston University Revision of Giles' Sunday School Examination A," *The Indiana Survey of Religious Education* (George H. Doran Company, 1925; now published by Institute of Social and Religious Research), Vol. II, ch. 15.
26. HARTSHORNE, H., "Measurements of Growth in Religion," *Religious Education*, 14: 148-155 (1919).
27. HARTSHORNE, H., and MAY, M. A., *Studies in Deceit* (The Macmillan Company, 1928).
28. HARTSHORNE, H., MAY, M. A., and SHUTTLEWORTH, F. K., *Studies in the Organization of Character* (The Macmillan Company, 1930).
29. HARTSHORNE, H., and MAY, M. A., "Testing the Knowledge of Right and Wrong," *The Religious Education Association Monographs*, No. 1 (July, 1927). Also in *Religious Education*, 21: 63-76, 239-252, 413-421, 539-554, 621-632 (1926); and 22: 523-532 (1927).
30. HEALY, W., and FERNALD, G. M., *Tests for Practical Mental Classification*, Psychological Monographs, Vol. XIII, 2, No. 54 (1911).
31. HIGHTOWER, P. R., *Biblical Information in Relation to Character and Conduct*, University of Iowa Studies in Character, Vol. III, No. 2 (1930).
32. JACOBSON-LASK, L., "Über die Fernald'sche Methode zur Prüfung des Sittlichen Fühlens und über ihre Weitere Ausgestaltung." *Beiheft zur Zsch. f. Angew. Psychologie*, 24, 1-84 (1920).
33. JOHNSON, G. A., "An Experimental Investigation of the Psy-

Tests of Conduct Knowledge and Judgment 297

- chology of Moral Judgment," *British Journal of Psychology*, 15: 255-265 (1925).
34. KOHS, S. C., "Ethical Discrimination Test," *Journal of Delinquency*, 7: 1-15 (1922).
35. LOWE, G. M., and SHIMBERG, M. E., "A Critique of the Fables as a Moral Judgment Test," *Journal of Applied Psychology*, 9: 53-59 (March, 1925).
36. QUADFASEL, F., "Die Methode Fernald-Jacobsohns, eine Methode zur Prüfung der moralischen Kritikfähigkeit-und nicht des sittliche Fühlens," *Arch. f. Psychiat. u. Nervenker.*, 74: 1-38 (1925).
37. RAUBENHEIMER, A. S., *An Experimental Study of Some Behavior Traits of the Potentially Delinquent Boy*, Psychological Monographs, Vol. XXXIV, 6, No. 159 (1923), 107 pp.
38. ROSNER, R., "Moral Prüfungen bei Jugendlichen," *Allg. Zsch. f. Psychiat.*, 86: 91-101 (1927).
39. SCHWESINGER, G. C., "Slang as an Indication of Character," *Journal of Applied Psychology*, 10: 245-263 (1926).
40. ———, *The Social-Ethical Significance of Vocabulary*, Teachers College Contributions to Education, No. 211 (1926).
41. SHARP, F. C., "An Objective Study of Some Moral Judgments," *American Journal of Psychology*, 9: 198-234 (1898).
42. ———, "The Study of Influence of Custom on Moral Judgment," *Bulletin of University of Wisconsin*, No. 236 (1908).
43. SLAVENS, G. S., and BROGAN, A. P., "Moral Judgments of High School Students," *International Journal of Ethics*, 38: 57-69 (1927).
44. SNYDER, A., and DUNLAP, K., "A Study of Moral Evaluations by Male and Female College Students," *Journal of Comparative Psychology*, 4: 289-324 (1924).
45. TANAKA, S., "A Study of Moral Evaluation by Boys and Girls," *Japanese Journal of Psychology*, 1: 93 (1926).
46. TERMAN, L. M., *Genetic Studies of Genius*, Vol. I (Stanford University Press, 1925), 648 pp.
47. WATSON, G. B., "Experiments with Religious Education Tests," *Program Papers*, No. 5 (Association Press, 1926).
48. WEBER, C. O., "Moral Judgment in Female Delinquents," *Journal of Applied Psychology*, 10: 89-91 (1926).
49. WELLS, F. L., *Mental Adjustments* (D. Appleton and Company, 1917).

Chapter IX

PERFORMANCE TESTS

THE most obvious method of measuring conduct is by means of the direct record of actual conduct itself. This method has had an extensive try-out in the psychological laboratory, and more recently there have been excellent field studies using these difficult techniques. The measurement of conduct presents certain difficulties that make it costly and exacting. In the first place conduct does not usually leave behind a record which can be studied at leisure. In the testing of ability, for instance, we can give an individual certain tasks to perform such that his performance of these tasks will itself provide a record which can be studied, evaluated, and utilized as a measuring stick. But to obtain a record of conduct necessitates considerable maneuvering and arrangement of the situation. Hartshorne and May, for instance, have given children school tests in arithmetic, not for the purpose of discovering their abilities in arithmetic but for determining whether the children would cheat when given the opportunity. Maller* also used tests in arithmetic in order to determine whether children prefer to work for themselves or for the group.

Observation may be used in the direct measurement of conduct, and certain techniques have been successfully worked out recently using observation methods. Ordinarily, however, this requires considerable skill and practice, and the method can never be as objective as when some record is left behind to be studied.

A second difficulty which the direct measurement of conduct presents is that of keeping the pupil unaware that he is being measured. In order to disarm the person being measured he must be thrown off the scent, as it were, by telling him to do one thing while at the same time giving him the opportunity to do something else. When one speaks of the measurement of conduct, he means measuring the number of times certain responses follow

* Maller, J. B., *Coöperation and Competition*, Teachers College Contribution to Education, No. 384 (1929).

certain antecedent situations. For instance, to measure cheating Hartshorne and May (10) counted both the number of times a child would cheat and the amount of cheating he did, when given the opportunity. But if they had told their subjects that they were watching to see if any one cheated, that would have been a new element in the situation, the children would have been on their guard, and the cheating presumably would have been much less. The direct measurement of conduct cannot issue from a frontal attack, as the measurement of ability can. It must catch the subject unawares.

May and Hartshorne * have constructed a complicated *schema* to describe this distinction between the measurement of conduct and the measurement of ability. They have classified situations into (a) *natural*, such as would ordinarily be met in the course of life's experiences whether the psychologist happened along or not (these *natural* situations are further divided into *uncontrolled* and *controlled*, the latter being planted so as to occur when and where the experimenter may observe the results) and (b) *experimental*, which are so artificial and unusual as never to occur except when set up by an investigator. They have also classified responses into (a) *natural* (either controlled or uncontrolled) and (b) *experimental*. These situations and responses are then paired in every possible grouping as follows:

Natural uncontrolled situation—natural undirected response

Natural uncontrolled situation—natural directed response

Natural uncontrolled situation—experimentally directed response

Natural controlled situation—natural undirected response

Natural controlled situation—natural directed response

Natural controlled situation—experimentally directed response

Experimentally controlled situation—natural undirected response

Experimentally controlled situation—natural directed response

Experimentally controlled situation—experimentally directed response

This *schema* is helpful in emphasizing the distinction between a test and the measurement of uncontrolled conduct. There seems to be some confusion, however, in trying to make a distinction between the *situation* and *controlling the response*. Certainly anything done to control the response must be part of the situation.

* May, M. A., and Hartshorne, H., "Objective Methods of Measuring Character," *Pedagogical Seminary*, 32: 45-67 (1925).

The distinction these authors are trying to make is between the *naturalness* of the situation (whether it occurs in the ordinary experience of the individual) and the directions given to the individual being measured. But these are both part of the situation. By the directions we determine whether we are testing ability or measuring conduct. In considering the naturalness of the situation we raise the question of the transfer value of the measurements.

This brings us to a third and fundamental difficulty in using actual performance as a measure of conduct. If transfer is general, and if there is high correlation between various exhibitions of the same trait; then the laboratory situation is as valuable as the uncontrolled natural situation for the measurement of conduct.

This point of view is illustrated by the following anecdote:

"A gentleman advertised for a boy, and nearly fifty applicants presented themselves. Out of that number he selected one and dismissed the rest. 'I should like to know,' said a friend, 'on what ground you selected that boy, who had not a single recommendation?' 'You are mistaken,' said the gentleman. 'He has a great many. He wiped his feet when he came in, and closed the door after him, showing that he was careful. He gave his seat instantly to that lame old man, showing that he was thoughtful. He took off his cap when he came in, and answered my questions promptly, showing that he was gentlemanly. He picked up a book which I had purposely placed on the floor, and placed it on the table, and he waited patiently for his turn instead of pushing and crowding, showing that he was orderly and honorable. When I talked to him, I noticed that his clothes were brushed, his hair in order. When he wrote his name, I noticed that his finger-nails were clean. Don't you call these things letters of recommendation?'"

If, on the other hand, the transfer is small so that the various performances going under the same trait name correlate only slightly with each other, the response to the laboratory situation cannot be used as a measure of the response to the natural situation. If this latter be the state of affairs (and we shall find that it is), then we must carry out our measurements in the exact situations which we think important. For then the response has no reference beyond the narrow situation in which it arose.

Let us not anticipate our conclusion here. Interest in a scientific attack on the problems of conduct is widespread, and the problem is a practical one. We are interested in selecting persons who will be successful in business and industry, or in advising

individuals so that they may make the most happy occupational adjustment. We are interested in measuring a person's adaptation to his surroundings or his ability to adapt himself so that if he is mentally unhealthy he may be given proper care and treatment. We are interested in discovering those who are apt to become enemies of society so that they may be reëducated or segregated before the trouble arises. In other words we are not so much interested in whether a child eats candy between meals when it is given to him, or cheats in a test in arithmetic when the teacher is out of the room and the results will affect his chances for promotion, or fails to continue doing his homework when a fire engine clangs past the house, as in whether he has healthy habits, or is honest, or is studious. If these specific responses to these specific stimuli are symptomatic to a slight degree only, we must acknowledge the facts and base our conclusions and practice accordingly.

In the following pages the outstanding attempts to measure conduct directly will be described. The techniques themselves will be described, and the method of scoring illustrated. Reliability figures and intercorrelations will be given when available. Finally, generalizations will be made as to the value and significance of the direct measurement of conduct.

Before going into detail on the various tests, let us first stop for a brief summary of the historical setting of the work on these tests. Although the earliest work on the dynamic factors of the personality, carried out strictly in the psychological laboratory, had only an abstract, theoretical interest, it set the scene for the later practical applications. The applied work on these performance tests of conduct has been carried out within the following areas of study:

1. The characteristics of delinquents.
2. The problems of cheating in school tests and examinations.
3. The value of certain educational programs (such as the Boy Scout movement) for the development of character.
4. The problems of persistence, caution, etc.
5. The character traits of *gifted* children
6. The development of young children.

These various fields of investigation have been more or less distinct, and only more recently have they tended to coalesce and to borrow techniques and findings from one another.

One outstanding piece of work was that of Voelker (17). His investigations were an outgrowth of the Indiana Survey of Religious Education (1) undertaken in 1920 and 1921. The collapse of the ill-fated Interchurch World Movement, which was born on the wave of enthusiasm following the World War, necessitated the curtailing of the original plans for the survey, and the results were salvaged by the Committee on Social and Religious Surveys of New York City. The committee in charge of the curriculum could not agree upon the advisability of using a set of "moral conduct tests" in the survey, and since both agreement and funds were lacking, the work was dropped. Interest in tests of this type, however, was maintained, and P. E. Voelker made them the subject of a doctor's dissertation at Columbia University. In this pioneering and notable work he had the assistance of the Boy Scout organization, the encouragement of W. S. Athearn and others of the Indiana Survey committee, the direct help and guidance of E. L. Thorndike, and statistical advice from T. L. Kelley.

Kelley carried the most promising ideas of this work with him to Stanford, and a year or so later we find V. M. Cady (4), working under the guidance of Terman and Kelley, studying the problem of incorrigibility, and using some of the tests originated by Voelker. Later Raubenheimer (15) at Stanford worked on a battery of tests which should predict tendencies toward delinquency.

Interest in the development and reorganization of religious education continued, however, and converging demands from various religious organizations finally led the Institute of Social and Religious Research to supply funds for the establishment of the Character Education Inquiry in 1923. Hugh Hartshorne and Mark A. May were selected as investigators, and the wisdom of the choice has been demonstrated by subsequent results. Originally planned to be a three-year study, starting in September, 1924, it was eventually extended another two years, till 1929.

The first year's program consisted of a study of tests of *honesty* and *trustworthiness*, extending and refining techniques originated by Voelker. These tests were followed by other groups of performance tests to measure *helpfulness*, *inhibition*, and *persistence*. At the same time tests of *moral knowledge* and *attitude* were studied. Finally the whole battery of tests which had been developed by the Character Education Inquiry was given to over 850 children, and a study of interrelationships and integration

was undertaken. The inclusiveness and high standards set by the two investigators has established the study of performance tests in a position unrivaled in other fields of educational research.

Tests of Honesty

These tests have been studied so extensively and reported so completely and ably in Hartshorne and May's *Studies in Deceit* and *Studies in Service and Self-Control* that they will receive somewhat summary treatment here. The reader is referred to the books themselves for a fuller explanation. In the first place, deceit is divided up into.

1. Cheating
2. Stealing
3. Lying

Test of cheating. 1. *Tests of copying.* Since much of the cheating in school is supposed to be copying from one's neighbor, it is natural that efforts should be made to discover the amount of this practice. Gundlach (9) tried the technique of seating students in pairs and giving some pairs identical test papers and other pairs dissimilar papers. When each member of a pair had the same paper, the number of errors was reduced (10 per cent) and the number of identical errors increased (18 to 20 per cent). But one does not know from this procedure how many or which individuals actually cheated. Hartshorne and May used test papers in which small changes were made in the questions which would be unnoticed except under close scrutiny, and tried to measure the amount of copying by noting the agreement between papers of pupils who sat side by side. However, they were not able by this method to differentiate accidental agreement from agreement due to copying. This might be done by noting the amount of agreement that exceeds what would be expected by chance, but such a method requires specially prepared test material.

2. *The duplicating technique.* The first form of this technique was devised by Voelker and called by him "Completion Test," although the name "Paraffin Paper Test" is more suggestive. Voelker's (18) description of the test is as follows:

This test is given on a prepared four-page folder, the sentences with the blanks being on page one, and the completed sentences

304 Diagnosing Personality and Conduct

being on page four. Page two is entirely blank. Page three has a coating of paraffine.

"Directions: The method of giving this test is as follows: A folder is placed before each subject, face side up. The subject is told that the completed sentences are on page 4 and for that reason he is not to look on page 4. The examiner remains in the room to see that those instructions are obeyed.

"When the time is up, the subject is requested to open the folder and to place it on the desk before him in such manner that he can see pages 1 and 4 at the same time. This procedure will lessen the chances of the subject's discovery of the paraffine on the inside of the folder, which contains a record of his effort to complete the test. The subject is requested to score his own paper, using page 4 as his model. During this part of the work the examiner absents himself to give the subject opportunity to cheat if he desires to do so.

"A comparison of the record made on the waxed surface, with the record as handed in on page 1 will reveal whether the subject attempted to cheat.

"Scoring: Score the subject 10 if he made no attempt to cheat. Score him 0 if he cheated."

Cady, who used this test, had the pupils score themselves both on the sum of errors made and for the number of spaces not filled, the latter being an additional stimulus to leave as few spaces unfilled as possible. Cady used the method of scoring for the presence or absence of cheating as described by Voelker (called *fact* score), and also used a score which represented the number of changes made or the amount of cheating (called *amount* score). "Graded scores do not correlate any more highly with the criterion than do the two interval scores."

Cady found the reliability of the paraffin-paper completion test of fifteen items to be .578, using only "fact" scores. Scoring for "amount" showed about the same reliability. This test correlated only .188 with the criterion of incorrigibility.

Hartshorne and May used a simpler "duplicating technique." After the questions of any kind of objective or short-answer test had been answered, the papers were collected and taken away, and a duplicate of each paper was made by clerks. At a later time when the papers were returned to the pupils together with an answer key, the pupils were instructed to score their papers, the examiner presenting all the while every opportunity to those who wished to change their answers to agree with the key. The test papers were then compared with the original duplicates and

CARDBOARD TEST

Wait for the signal for each trial. Put the point of your pencil on the cross at the foot of the oval. Then when the signal is given shut your eyes and put the figure 1 in each circle, taking them in order. For the second trial, put the figure 2 in each circle. For the third trial, the figure 3, etc. After each trial put a check mark in the score box under the number of each circle you succeeded in striking, count the checks and enter the total in the column headed T at the right of the score box. After the last trial add up Column T. This is your score. The maximum score is 50.

S C O R E

Circles	1	2	3	4	5	6	7	8	9	10	T
Trial 1											
" 2											
" 3											
" 4											
" 5											
Total											

From Hartshorne and May, *Studies in Deceit* (The Macmillan Company, 1928). By permission of the Macmillan Company, publishers.

306 Diagnosing Personality and Conduct

changes noted. This simple and effective device was expensive and time-consuming, however, and has not been used extensively.

3. *The improbable achievement technique.* The earliest form of this technique was devised by Voelker, who called it the "Card-board Test." Both Cady and Hartshorne and May used the same device. The figure on page 305 shows the form used by the Character Education Inquiry. The directions indicate the method.

All three experimenters tried out the possibility of completing the test honestly. Voelker, using five circles, reports, "This feat is impossible of accomplishment as far as it has been possible to determine. The examiner himself, after long practice and fifty successive tests, was unable to score a single success. No trustworthy subject has been found who is able to accomplish this feat." Cady says, "The chances of obtaining a perfect score in the 'circles' is very remote. One would be required to hit each circle, large or small, in every one of the five trials." Ten adults tried to do this, each making ten trials. In no instance, despite the practice acquired, were both circles 3 and 4 hit by the same person in his ten trials. Hartshorne and May verified this with school-children under conditions where there was no possibility of cheating.

Similarly Cady and also Hartshorne and May used "nested squares" and "mazes" as coördination tests. In these tests the pupil was instructed to draw his pencil in the pathway between squares or through mazes without touching the sides, tasks which are impossible with both eyes shut. Cady was careful in his forms to have the first exercise so easy that there was a possibility of performing it successfully. "The situation is one in which honesty on the part of the subject appears to him to be taken for granted. Integrity of purpose is implied in the care with which the experiment is carried out."

Voelker merely gave a score of 10 or 0 according to whether the subject gave evidence of having cheated or not. Hartshorne and May worked out both "amount" of cheating and "fact" of cheating scores for all their tests. In these particular tests the amount score was the number of corners correctly turned in the squares and mazes and the number of figures correctly placed in the circles. The fact score (cheating) was 14 or more corners in the

squares, 13 or more marks in the circles, and 32 or more corners in the maze.

Cady (4) finds that his test of "circles" correlates with his "maze" test .744, and this may be used as a measure of reliability. With forty-four cases the intercorrelations are:

TABLE 59
INTERCORRELATIONS OF CIRCLES, SPACES, AND MAZES TESTS
(From Cady, 4, p. 57)

Circles and spaces.....	.592
Circles and mazes.....	.513
Spaces and mazes.....	.630

Each of the three correlates with the criterion of incorrigibility as follows:

TABLE 60
CORRELATIONS OF CIRCLES, SPACES, AND MAZES TESTS WITH INCORRIGIBILITY
(from Cady, 4, pp. 54, 57)

	150 cases	44 cases
Circles318	.526
Spaces297	.324
Mazes524
Circles plus squares398	...

Hartshorne and May (10 II, p. 97) find the reliability for their squares, circles, and mazes tests based on intercorrelations (predicted by Spearman-Brown formula) to be .721, and based on retests to be .750.

Hartshorne and May also used the "improbable achievement" technique in the following forms: (a) puzzle peg, (b) the fifteen puzzle, and (c) weight discrimination test.

To illustrate the method, the last form only will be briefly described. Seven small pill-boxes were filled with cotton batting and buck-shot so that their weight in grams was (a) 3.6, (b) 3.7, (c) 3.8, (d) 3.9, (e) 4.0, (f) 4.1, (g) 4.2. The differences between adjoining boxes were so slight that they could not be detected, making it impossible to arrange them in order of weight except by chance. Numbers from one to seven were written on the bottom of the boxes.

"The instructions were to turn the numbers down and arrange the boxes in the order of their weight. After the first trial the

pupils were told to look at the numbers on the bottom and copy* these numbers off on the score sheet to show how they had been arranged. They were then told that the correct arrangement was the serial order 1, 2, 3, 4, 5, 6, 7, and were asked to turn the numbers down again and not look at them during the second trial." (10, p. 59.) *

Amount of cheating was determined by counting the number of boxes in correct order in any one trial. A table was drawn up for determining the amount of cheating when combining the scores on the two trials which were given. The fact of cheating was credited if a score of seven (all seven boxes in correct position) was made in the second trial; or a score of five on the second trial, provided the score was five or seven on the first trial.

The reliability of this test as determined by intercorrelations is .750.

4. *Double testing technique.* The overstatement test is one of the most widely used tests employing the double testing technique. It was first devised by Voelker, developed by Cady and Raubenheimer and Ruch, and later studied by Cushing and Ruch (7), and Woodrow and Bemmels (20).

Raubenheimer used it as follows (15, p. 39):

First there was a practice form, introduced in order to motivate the desire to make overstatements. (Voelker even offered a prize of a box of candy to the boy making the highest score.)

The directions read:

How many of the 10 things mentioned below can you do? We want you to mark yourself on each one in this way:

I can do it very well	2 points
I can do it fairly well	1 point
I know nothing about it	0

Now go through the list and give yourself a grade of two points on each one that you can do very well; give yourself one point on each one that you can do fairly well; and give yourself 0 on each one that you know nothing about at all.

A perfect score is 20 points. We want to see how many of you can get a perfect score. After you finish we will have you stand up and tell your score.

* From Hartshorne, H., and May, M. A., *Studies in Deceit* (1928). By permission of The Macmillan Company, publishers.

Performance Tests

309

Begin here.

	<i>Score</i>
1. Can you swim?points
2. Can you skate on roller-skates?points
3. Can you drive a car?points
4. Can you ride on horseback?points
and so on for ten questions.	

Then came the test itself. The directions are similar to those used in the practice sheet. Each pupil is asked to rate himself 2, 1, or 0, according to his confidence in ability to answer the questions. Some of the questions are (15, p. 40):

	<i>Points</i>
1. Do you know who discovered America?
2. Do you know who saved the life of Captain John Smith?
3. Do you know much about the Boy Scouts?
4. Do you know where the Pyramids are?
5. Do you know how to play checkers?
and so on for forty questions.	

Finally a test is given to determine whether the pupil is really able to answer the questions which he claims to be able to answer. The test is in the conventional multiple-choice form.

The first five questions are (15, p. 44):

1. America was discovered by Drake, Columbus, Balboa, Cook.
2. The Indian girl who saved Captain John Smith was Pocahontas, Uncas, Hiawatha, Minnehaha.
3. A Boy Scout must not eat candy, ride horseback, chew gum, smoke cigarettes.
4. The pyramids are in Arabia, Palestine, India, Egypt.
5. The king-row is used in checkers, cards, dominoes, croquet.

The score on this test is expressed in percentages of overstatement or understatement. This is calculated in the following manner:

Score on Part I—the points which the subject gave himself.

Score on Part II—(the rights minus $\frac{1}{3}$ the wrongs) times 2.

Final Score—the per cent that the score on Part II is less or more than the score on Part I.

Thus: A subject gave himself 50 points on Part I. He had 20 items correct; 6 were wrong on Part II. His score on Part II was 20 minus $\frac{1}{3}$ of 6, which is 18; multiplied by 2, viz., 36. His final score, therefore, was overstatement to the extent of 28 per cent.

310 Diagnosing Personality and Conduct

Cady reports a correlation of .505 between the estimate of ability and estimate of knowledge and gives the reliability as .579.

Raubenheimer found a reliability of .76 for his overstatement test, which is raised to .86 by the Spearman-Brown formula for the two forms combined. Terman reports a reliability of .78 for Raubenheimer's Overstatement Test.

Cady found his overstatement test to correlate .414 with the criterion of incorrigibility. Raubenheimer found biserial r 's of .63, .47, .48, and .32 between the most stable individuals in normal school groups and delinquent boys in a parental school. Cushing and Ruch found a biserial r of .13 between normal and delinquent girls. Woodrow and Bemmels found correlations of .56 and .43 between character ratings and the results of an overstatement test for children in kindergarten and a nursery school, respectively. Evidently this test has symptomatic value in differentiating between children who adjust themselves well and poorly to the social life of the school.

A variation of this Overstatement Test known as the "Books Read Test" is so similar to it that it will be described here, although it does not in reality use the double-testing technique. This test, which is ascribed by Terman to Knight, claimed by Ruch as originating with himself, and credited by Raubenheimer to Franzen (8), was first published by Raubenheimer. The test consists of a number of book titles, some of which are fictitious. The pupil is to check those which he has read. The score is the number of fictitious titles checked, "it being thought that such misstatements on the part of the subject might be an index to his mental honesty." Hartshorne and May criticize this test on the ground that it is impossible to differentiate between "errors of honest recognition" and dishonesty. Raubenheimer reports, however, that the test which he used with ten fictitious titles in each form showed reliability of .59 raised by the Spearman-Brown formula to .74 for both forms together. The test shows biserial r 's of .41, .31, .45, and .37 in differentiating between stable boys in normal groups and delinquent boys in a parental school.

The most successful form of the double-testing technique, devised by Hartshorne and May, is the technique for measuring cheating which they have used most extensively. In brief it consists of two equivalent forms of the same test, one given under conditions where no cheating is possible, and the other under

conditions where cheating is possible. Usually this has been accomplished by passing out an answer key with each paper the first day, telling each pupil to use it to score his paper with, but to keep it out of sight until he is ready to score; and then giving the equivalent form of the test without answer sheets on the second day.

Given two forms of the test equivalent in difficulty at all levels of the scale, it is possible to estimate statistically exactly how much variation between scores on the first and second tests there may be without cheating. Then if the difference between the first and second tests exceeds this figure, cheating may be suspected or even asserted as a fact.

For instance, in one of the arithmetic problem tests which were used, it was found that on repeating the test under conditions such that there was no possibility of cheating either time there was an average gain of 1.06 points (which we call practice effect), but that there was a variability of gains (or losses) which amounted to 3.10 points (standard deviation). The chances are less than 3 out of 1,000 that an error as large as 30 or 9.3 points will be made. But since the practice effect is 1.06, we may expect the second score to be lower than the first score by 1.1 minus 9.3 or 8.2 points only 3 out of every 1,000 times. Arbitrarily assuming this chance to be negligible, Hartshorne and May decided that when the second score was 8 points or more lower than the first score on the arithmetic problem test, there was cheating. This they took as their "fact" score. They arbitrarily took a deviation of 2.92 standard deviations as the dividing line between the certainty of cheating and not cheating and used this unit as the "amount" score.

Two types of material were used with this technique. One type of test showed very little practice effect and was dependent hardly at all on speed. Four of these difficulty tests were used as follows:

1. Arithmetic problems
2. Completion sentences
3. Information test items
4. Vocabulary test items

Speed tests were also employed using the same technique.

1. Test of adding two-digit combinations
2. Number checking
3. Cancellation of A's

4. Digit symbol substitution test
5. Making dots in small squares
6. Cancellation of a single digit

In these speed tests cheating may be accomplished by working beyond the time limit, so that no key is needed. The advantages of using these speed tests are that they require less time, and since they are so unlike ordinary school work there is less overstimulation to cheat. On the other hand, they require more skill to administer, and there is a larger practice effect. This latter is partly controlled by giving an additional practice form before cheating is permitted.

The reliabilities of such honesty tests are not high. Using the power tests Hartshorne and May found the following reliabilities (10, II, p. 93):

TABLE 61
RELIABILITY OF HONESTY DIFFICULTY TESTS*
(from Hartshorne and May)

<i>Honesty tests</i>	<i>Reliability coefficient</i>
Arithmetic problems484
Completion sentences485
Information535
<hr/>	
School tests752
Vocabulary	
Home test205

The intercorrelations are: arithmetic and completion, .454; arithmetic and information, .481; and information and completion, .450.

The reliabilities of the honesty speed tests are (10, II, p. 95):

TABLE 62
RELIABILITY OF HONESTY SPEED TESTS**

Additions424
Number checking322
Cancellation of A's502
Digit symbols365
Dots397
Cancellation of digits490

* From Hartshorne, H., and May, M. A., *Studies in Deceit*, 1928. By permission of The Macmillan Company, publishers.

** *Ibid.*

Besides these techniques for measuring cheating in school situations, Hartshorne and May used other forms of testing to test cheating at home, in athletic contests, and in parlor games. The test of cheating at home consisted simply of giving one of the tests already described, to be done at home. Results were compared later with a test taken in school under standard conditions.

The tests in athletic performance consisted of standard strength or performance tests in which two trials were given and an opportunity to cheat was permitted on one of the tests. The tests used were:

1. The hand dynamometer for measuring strength of grip
2. The spirometer for measuring lung capacity
3. Chinning
4. Standing broad jump

As in the school tests in which the duplicate testing technique was employed, the possible difference between two trials in which there was no cheating was determined, and any difference which exceeded this when cheating was allowed was called cheating. The limits beyond which any difference was called cheating are as follows:

Dynamometer	3 kilograms
Spirometer	25 cubic inches
Chinning	3 times
Broad jump	7 inches

The reliability of these tests determined by intercorrelation and raised by the Spearman-Brown formula is .772.

Games ordinarily played at children's parties were also used as situations in which cheating might take place and where it could be measured. The games used were:

1. A peeping stunt
2. Pinning the tail on the donkey or the arrow on the target
3. Bean Relay
4. The Mystery Man

Only one of these, the Bean Relay, will be described here to illustrate the method. This is a modified potato race, beans being used instead of potatoes. The races were run off five at a time to simulate a contest. Besides the home box in which the beans were to be deposited there were three other boxes, the first and second with three beans each, and the third with ten or more beans. The

314 Diagnosing Personality and Conduct

object was to carry as many beans as possible one at a time to the home box in thirty seconds. Observers counted the number of trips each child made. When the heat was over, the number of beans in the "home box" was also counted and if there were more beans than runs this was *prima facie* evidence of cheating.

The reliability of these tests is not reported.

Tests of stealing. Four tests were used by Hartshorne and May to measure stealing. These are:

1. The Planted Dime Test
2. The Magic Square Test
3. The Coin-Counting Test
4. The Mystery Man (one of the party games)

In the coin-counting test (which will be described to illustrate the method) each child was given a mimeographed sheet containing such problems * as (10, p. 93):

What three (3) coins add up to forty (40) cents?

() quarters () dimes () nickels () pennies

What three (3) coins add up to twelve (12) cents?

() quarters () dimes () nickels () pennies

What three (3) coins add up to thirty-one (31) cents?

() quarters () dimes () nickels () pennies

Boxes were also distributed containing the following coins: 1 quarter, 4 dimes, 4 nickels, and 4 pennies. "The pupils were told that it was a money counting test and that, in order to make it a real test, the coins were to be used to count with instead of writing on the paper." When they found the right combination with the coins they entered the number of coins used on the mimeographed sheet. These sheets were collected after a certain time limit and then the directions were given: "Now put all the money back in the box, put the band around it, and we will collect it. Pass the boxes to the center aisles." Opportunity was afforded for any child to take some of the money if he was so inclined.

To identify the boxes the following device was used. A blurred space was purposely left in one of the problems and the children were instructed to copy the number on the bottom of their box in this space. Since they signed their names to the sheets, identification of the boxes was effected.

The reliability of these tests has not been stated.

* From Hartshorne, H., and May, M. A., *Studies in Deceit* (1928). By permission of The Macmillan Company, publishers.

Tests of lying. Two tests of lying were constructed, both paper-and-pencil tests, one to measure lying to avoid disapproval, the other to measure lying to gain approval.

One of these tests to measure lying to avoid disapproval was a set of questions of which the following are samples * (10, p. 95):

33. Did you ever cheat on any sort of test?
34. Have you ever cheated on such tests more than once?
45. On some of these tests you had a key to correct your paper by.
Did you copy any answer from the key?

Two scores were given. One was a truthfulness index, which was determined by the number of admissions of having cheated on the tests of cheating previously administered. The other was an index of lying which was obtained by comparing the answers with the actual results on the cheating tests. The scoring method is somewhat complicated but involves a comparison of the answers to the questions with actual conduct on the previous test.

The other test designed to measure *lying to win approval* consisted of a number of questions concerning "specific acts of conduct which on the whole have rather widespread social approval, but which at the same time are rarely done." Examples are ** (10, p. 98):

- | | |
|--|--------|
| 4. Do you usually report the number of a car you see speeding? | Yes No |
| 5. Do you always preserve order when the teacher is out of the room? | Yes No |
| 13. Do you usually pick up broken glass in the street? | Yes No |
| 29. Do you read the Bible every day? | Yes No |

To determine the critical point that divides honesty from lying the investigators had a graduate class answer the questions in a way that would represent their childhood. Then a point differing by three standard deviations from the mean of each test for this adult group was taken as being the critical point a score above which indicated the presence of lying.

Another type of deceit which seems to be neither cheating, stealing, nor lying is that of failing to report what is credited to one in error. This partakes of all three—lying, stealing, and cheating. Miller (12) tested this by planting errors in the scores on

* From Hartshorne, H., and May, M. A., *Studies in Deceit* (1928). By permission of The Macmillan Company, publishers.

** *Ibid.*

papers returned to students enrolled in the summer session of a large university. The paper of every other student was correct, so that there would be little likelihood of arousing suspicion. Students were asked to check the marks which they received. The results indicated that fewer students reported the error in scores which were too high than in scores which were too low. This technique does not permit fixing the guilt on any individual student, as it is impossible to separate intention from carelessness or negligence.

All in all, the tests for measuring deceit devised by the Character Education Inquiry provide a very comprehensive means for surveying the various types of deceitful behavior in several different situations. So thoroughly has the work been done that the techniques will probably not have to be added to for many years. When the thirty-two different tests proposed are reviewed, it is seen that their administration is expensive, both in time and in money. For this reason alone if for no other, the use of tests of this kind will probably always be restricted to experimental work where adequate funds are available.

All of the results reported by Hartshorne and May in validating their tests cannot be repeated here. Three tables will be given showing the reliability of types of deceptive behavior studied, the intercorrelation between types of deceptive behavior, and average correlation between single tests of different techniques.*

TABLE 63
RELIABILITY OF TYPES OF DECEPTIVE BEHAVIOR
(from Hartshorne and May, 10, II, p. 122)

<i>Types of deception</i>	<i>Reliability</i>
I. The classroom type of cheating	
A. Copying from a key	.871
B. Adding more scores after time is called	.825
C. Peeping	.721
D. Faking the solution to a puzzle	.750
II. The out-of-classroom-types of cheating	
E. On home work	.240
F. On athletic contests, faking a record	.772
G. In parties: faking, peeping, stealing	Not known
III. H. The stealing type of dishonesty	Not known
IV. The lying type of dishonesty	
I. Lying for approval	.836
J. Lying to escape disapproval	Not known

* From Hartshorne, H., and May, M. A., *Studies in Deceit* (1928). By permission of The Macmillan Company, publishers.

Performance Tests

317

TABLE 64

INTERCORRELATIONS OF NINE TYPES OF DECEPTIVE BEHAVIOR

The measures of any type represent the composite of all tests used.

(from Hartshorne and May, 10, II, p. 123)

	B	C	D	E	F	G	H	I
A	.450	.400	.400	.172	.288	.118	.143	.350
B		.374	.425	.193	.345	.169	.173	.248
C			.300	.234	.100	.250	.200	.108
D				X	.300	.122	.346	.256
E					.142	-.015	-.010	.400
F						.118	.283	.230
G							.210	-.004
H								.132

TABLE 65

AVERAGE INTERCORRELATIONS BETWEEN SINGLE TESTS OF DIFFERENT TECHNIQUES

(from Hartshorne and May, 10, II, p. 212)

	B	C	D	F	H	I
A	.292	.285	.291	.198	.127	.312
B		.219	.255	.194	.128	.254
C			.196	.062	.160	.161
D				.184	.283	.208
F					.162	-.003
H						.132

A progressive lowering in reliability may be seen as one goes from a situation that is very similar to one that is different. For instance, the average correlations with a single test of copying from a key are as follows:

One test of copying from a key with another test of copying	.696
One test of copying from a key with one test of adding more scores after time is called (both school tests).....	.292
One test of copying from a key with one contest (the latter being outside the class-room).....	.198

As the situation presents features which differ the response varies, or to quote Hartshorne and May, "an individual behaves similarly in different situations in proportion as these situations are alike." The conclusion from this is that a single test of deceit has little symptomatic value of deceit in general. The value of a single test as an honesty test depends on the degree to which the single situation has features which are common to all the situations in which honesty may be displayed. In applying this rule one must remember that few individuals respond to the abstract

318 Diagnosing Personality and Conduct

features of a situation—to the honesty factor, for instance—but to such concrete things as the teacher, pencil, paper, school marks, coins, stamps, tickets, social stimulation, ease with which deceit may be accomplished, etc. A test which employs one set of these—paper, pencil, class-room, teacher out of the room, pupils taking a test, knowledge that the test counts for promotion—may be quite different so far as incentives to dishonesty go from a situation that involves a contest, spoons, boxes, beans, running, observers with pads and pencils. We should expect a test of cheating in one situation to give results similar to a test in the other situation only to the degree that pupils recognize (respond to) some element common to both.

Tests of Suggestibility

The precise bearing of "suggestibility" on conduct is uncertain. However, the matter is worthy of investigation, if only to point out the negative effects of suggestibility. Whipple, in his manual of *Mental and Physical Tests*, summarized in his usual complete fashion the work that had been done with tests of suggestibility up to the time that he wrote (1915). He describes these tests as follows (29, p. 222):

"In them (tests of suggestibility) the experimenter seeks, by suitable arrangement of the test-material or of the instructions, to induce the subject to judge otherwise than he naturally would—to induce him, for example, to judge equal lines or equal weights to be unequal, or to perceive warmth where there is no warmth, etc. If the attempt is successful, the subject is said to have 'yielded,' or to have 'accepted' the suggestion; if unsuccessful, he is said to have 'resisted' the suggestion. The degree of his suggestibility is indicated by the quickness or frequency of his 'yields.'"

Whipple describes five tests:

Suggestion by the Size-Weight Illusion

Suggestion by Progressive Weights

Suggestion by Progressive Lines

Suggestion by Line-Lengths by Personal Influence

Suggestion by Illusion of Warmth

The most work has been done on the first of these tests, and hence it will be described briefly here. Its purpose is to see to what

degree big things are judged to be heavier than small things. Two standard cylindrical blocks and twenty comparison blocks are placed before the subject.

"Both standards weigh 55 grams: both are 28 mm. thick, but the larger is 82 and the smaller 22 mm. in diameter. The 20 comparison blocks are all 28 mm. thick and 35 mm. in diameter, but the weights range from 5 to 100 g. by 5 g. increments.

"Place before the subject the larger standard block and say: 'Here is a block. I want you to find a block in this series of 20 blocks that seems to you just as heavy as this one. Lift it by picking it up edgewise with your thumb and finger like this (illustrate). Then try the first of these weights (at the left). If that doesn't suit, try the next, then the third, and so on till you find a block that seems equal to this one. Each time you must lift this block first, then the one you are trying in the series. Keep your eyes constantly directed at the weight you are lifting.' When the subject has selected an equivalent weight, the same procedure is followed with the second, or smaller, standard block."

The score is the difference between the weights of the standard blocks and the ones selected to be of equivalent weight. Whipple reports Gilbert to the effect that the illusion is well developed at the age of six years, apparently increases until age nine, and then decreases with advancing age. Various investigators have found that defective children with mental ages below four fail to get the illusion at all. Reports differ as to the relative suggestibility of the sexes. Practice seems, if anything, to increase the illusion.

Readers are referred to Whipple's manual for a description of the remaining tests.

W. Brown, working later, made very extensive researches on suggestibility, using a wide variety of situations including a "least perceptible" (imagined) sensation; the perception of change; a series of progressive increases in weights and lines; memory, recognition, and imagination; normal illusions; estimations of magnitude; simple esthetic preferences in the matter of properties in geometrical figures; and simple esthetic preferences in the quality of sensation.

His conclusions are (23, p. 424): "There are no individual differences which are sufficiently conspicuous to justify the experimenter in calling one person 'very suggestible' and another 'not suggestible.' There are no individuals who have consistently high or consistently low indices of suggestibility through a series of

tests. On the contrary, the experimenter is struck by the fact that the most skeptical individual will yield at times with surprising readiness to the suggestion, while a person who has yielded to some tests with very little apparent resistance will unexpectedly become very recalcitrant." This statement denying that there are tendencies to be *highly* suggestible or the contrary shall be tempered by a more exact statement as to the degree of intercorrelation. Brown found 472 intercorrelates between the different tests which he used, the average of which was $+.143$. This leads him to make the statement (23, p. 430): "It is probable that an individual who is more suggestible than another in one of these tests will prove more suggestible in another. Yet the actual amount of the correlation is so small and the number of negative instances so great that the 'probability' in the above statement can only be very slight. While it seems to be true that suggestibility is a trait more conspicuously developed in some individuals than in others, yet the individual differences are small and seem to be subject to reversal under the influence of conditions which are not within the control of the experimenter."

Prideau (27), writing in 1920, states from clinical experience, "(a) Suggestibility varies in different persons irrespective of the nature of the suggestion and of the suggestor. (b) Suggestibility varies in the same person at different times and under different conditions. (c) Suggestibility may have reference to a particular system of ideas only. (d) A person may be suggestible toward one person and not towards another."

In other words, one test of suggestibility is practically worthless because its results are so very specific. In order to test an individual for suggestibility, many tests must be used representing a variety of situations. But this would add to the time, difficulty, and cost of the testing and would defeat its own purpose.

M. Otis (26) attacked the problem later (1924) with a fresh approach. Criticizing previous work on the ground that various types of reactions had been grouped under the name *suggestibility*, she essayed to study "ability to resist a suggestion" by means of a group paper-and-pencil test. The test comes in two forms, A and B, of forty items each. Only twenty-two in each form may be used to measure suggestibility, however, the remaining items being innocent items from the Woodworth-Wells Directions Tests. Each

pupil is given a sheet on which he records his response, and the directions are read aloud by the examiner.

Sample item (26, p. 68):

*Directions read aloud by
teacher*

*Sheet on which pupil writes
his answers*

1. You see here some circles.
Write a word in the third
circle. (10")

1.



3. Write the answer to this
question: Do butterflies eat
bugs or green leaves? (10")

3.

5. Here are two squares and
two circles. If the circles are
larger than the squares, put
a dot in the center of the
smaller square. (Pause.)
When through with that,
write the letter *a* under the
larger square.

5.



Otis found correlations of sufficient size (average .56) between the two forms of her test to justify her in claiming "There is a *trait* that we may call ability to resist suggestion." Part of this reliability, however, comes via the influence of intelligence, which correlated with the tests .77, .72, and .75 on three different occasions. When a group of 100 children having the same mental age (9.0 to 9.11) were tested, the correlation between Form A and Form B dropped to .19. So we are forced to dissent from her conclusion that her test indicates the presence of a definite trait, "ability to resist suggestion," which exists apart from the intelligence involved in the situation.

Tests of Persistence

Tests of persistence have been few, and yet the scanty evidence which is available indicates that there is something in tests of this type which holds considerable promise for the diagnosis of conduct.

In school work there is a good reason to believe that persistence, or sticking to a task, is one of the main factors that helps to supplement or compensate for ability. Many a dull child has won a "passing" grade through the exercise of unusual diligence. There is

322 Diagnosing Personality and Conduct

also good evidence that some bright children tend to let go of a task too quickly when they have done enough to excel and to win the praise of the teacher and the approbation of fellow-pupils.

Persistence is also a prime factor of success in the workaday world. The rolling stone has won proverbial fame for failure. Studies of successful men, whether in business or in professional life, indicate that in every case there is a certain persistence of activity that produces the fruits. Here again nothing can completely compensate for lack of talent, but there is considerable leeway in the ebb and flow of affairs where sticking to it will cause one man to succeed where another man of equal ability but with less persistence would fail. Furthermore, there is evidence, as will be seen in the following pages, that persistence is also a factor in adequate social adjustment.

In 1911 Fernald (32, p. 331) reported at a meeting of the American Psychological Association a paper entitled "A Kinetic Will Test." He was interested in the problem of testing and examining delinquents and in discovering tests which would differentiate between them and normal youths. He used a strictly deductive approach in constructing his test. "It is essential to include in any comprehensive group of psychological tests to be applied in the classification of defectives among prisoners an adequate test for the function called *will*, since the success or failure of individuals depends so largely on the ability to endure and to continue to strive for the sake of achievement and in spite of fatigue and discouragement."

The test makes use of an apparatus which consists of a platform on which the subject rests his heels and a dial at the height of the eyes informing the subject of the height of the heels from the floor. The test is to see how long one can stand on tiptoes with heels an eighth of an inch or more from the floor. When the subject brings his heels to the floor, a bell rings indicating that the test is through. The dial also has a dead-line, and if the subject gradually weakens, the directions warn him that he must bring his heels to the floor when the little black disc passes the mark three times.

Dr. Fernald gave the test to 116 reformatory prisoners and twelve members of the senior class of the Ringe Manual Training School of Cambridge, Massachusetts, with the following remarkable results:

	<i>Median time</i>	<i>Lowest score</i>	<i>Highest score</i>
Reformatory group . . .	14' 54"	2' 30"	52' 45"
Normal group	36'	12'	2° 30' 6"

Comments on the result of the test show that in every case the decision to "give in" preceded the necessity of it; that varying degrees of weight and varying degrees of strength neutralize each other; that previous training was not a factor in differences with respect to the particular function tested. Fernald mentions as a point in favor of the test that it is independent of language, but in view of the doubtful value of other non-language tests for measuring conduct, this point must be taken with reservations.

Bronner (30) reports that in later experimentation with the Fernald test the norm was 50 minutes, making it inconveniently large for experimentation. She tried out another persistence test as follows: "The subject was given a pair of iron dumb-bells, each of which weighed two pounds. She was told on a given signal to take one in each hand and extend the arms level with the shoulders, holding the dumb-bell in a horizontal position. As soon as the arms were dropped about five inches or more, the time score was taken." In trying out this test Bronner found that but three out of twenty-six members of a group of delinquent girls reached or exceeded the median of a group of college students. "It would seem that the members of Group C (the college group) were much more willing to endure physical discomfort for the sake of a good record than were the members of Group D (the delinquent group). Very frequently girls in the latter group would remark, 'Oh, it hurts!' and drop the dumb-bells. They seemed on the whole to have much less will power and physical endurance, at least in matters where there was no necessity for continued discomfort other than mere pride in a deed well accomplished." Both of these tests show such remarkable differences between normal and delinquent groups that tests of this type are worthy of further trial. However, in future work care should be taken that the results do not depend on physical strength or special skills, but merely on willingness to endure.

Morgan and Hull (34) has introduced the use of a partially concealed maze which is impossible of accomplishment as a test of persistence. Not satisfied with a mere measure of *time*, they had observers rate the persistence of subjects on a nine-point

scale. They decided that the reliability shown by the judges in this subjective evaluation of persistence showed close enough agreement to warrant further experimentation with its use.

Persistence in taking mental tests where there is no time limit imposed is a natural method of testing persistence. Persistence has been suggested as a factor in the three-hour Thorndike "Intelligence Test for High School Graduates," also in the Thorndike "CAVD" tests, but the effect of persistence has never been disentangled for separate study from the factor of ability. Chapman made a preliminary study of persistence in a test of *word-building* scored by total number of words obtained and total time consumed. He found a considerable positive correlation (average .65) between persistence and success. Although this correlation is high, Chapman cites evidence to support the belief that it could be even higher if the capable children did not give up so easily. "It will be seen that whereas there are very great differences in speed of thought between the slow and rapid groups, the persistence of the slow group goes a long way towards compensating for the lack of speed."

Chapman (31) points out that persistence is extremely sensitive to group stimulation. Whereas in certain public schools the median time of work was only ten minutes, in a certain parochial school at the end of twenty minutes only three pupils out of forty had abandoned the task. He mentions four factors which may influence tests of this type, (a) habits of the pupils, (b) directions given by and prestige of the examiner, (c) setting of the test, (d) manner in which the children weight the importance of the result.

Very ingenious tests were devised by the Character Education Inquiry (33) to measure persistence. In the "Cross Test," the familiar and exceedingly difficult "Japanese Cross" puzzle, consisting of wooden sticks which can be made to interlock in the form of a cross, was used. With this was used a puzzle called the "Chinese Rings." Each child being tested was allowed to choose which one he would work with first, and the score was the time spent working on the first puzzle tried. Magic number squares and magic word squares were used as puzzles in the same way.

A different type of measure of persistence was obtained from monotonous tests of simple additions. These tests were also used in the measurement of coöperation. Twelve sheets were added in succession, requiring work for twenty-four minutes without rest.

Six of the sheets counted for each individual and six for the class. The persistence score was the difference between the scores made on the first two sheets and on the last two sheets.

The reliabilities of these tests were fairly high. They were obtained by correlating the first half of the testing with the second half, and then raised by the Spearman-Brown formula to indicate what reliability could be expected from the complete testing.

TABLE 66
RELIABILITIES OF PERSISTENCE TESTS *

	Reliability coefficient	Reliability corrected by Spearman-Brown formula
Story resistance59	.75
Cross and ring with magic number square38	.55
Persistence for self79	.88
Persistence for class85	.92
Whole battery80	.89

But the intercorrelations show the usual tendency to be very low.

TABLE 67
INTERCORRELATIONS OF PERSISTENCE TESTS **

	2	3	4	5
1. Story resistance35	.16	.15	.18
2. Magic number squares38	.20	.15
3. Cross and ring22	.18
4. Persistence for self42
5. Persistence for class				
Average inter- r239			

Coöperation or Service

Original tests of coöperation for use in the class-room were devised by the Character Education Inquiry. Three of these will be briefly described. For a complete description, the reader may turn to Hartshorne, May, and Maller's *Studies in Service and Self-Control*.†

* From Hartshorne, H., May, M. A., and Maller, J. B., *Studies in Service and Self-Control* (1929), p. 330. By permission of The Macmillan Company, publishers.

** *Ibid.*

† The Macmillan Company, 1929.

326 Diagnosing Personality and Conduct

In the "Kits Test" there were given to each child school pencil-boxes containing ten articles: drinking cup, pencil sharpener, ruler, eraser, pen, penholder, double pencil, and three other pencils. Each child was allowed to give away any number of the articles in his kit that he desired to help make up kits for poor children who had none. A table was drawn up weighting the value of the articles in inverse proportion to the percentage given away, and pupils were scored according to this key on the articles they gave away.

In the "Envelopes Test" four envelopes were distributed to the children and directions were given to find at home and put in the envelopes jokes, puzzle pictures, short stories, or a beautiful picture if they wanted to help contribute these to children in hospitals. Again a score sheet was prepared giving values for each envelope according to the number of items supplied. The score was the sum of the credits thus earned.

In the "Money Voting Test" a ballot was prepared for each child in a class to vote on how he would like to see prize money disposed of. The numbers in brackets are the rank order of the items in social significance according to the judgments of the experimenters.

- (4) Give all the money to the boy or girl scoring highest on the test.
- (2) Buy something for our school, such as bats, balls, skipping-ropes, a big picture.
- (3) Buy something for the room, such as a picture, a globe of goldfish, some plants.
- (5) Divide the money equally among the members of the class.
- (1) Buy something for some hospital child or some family needing help or for some other philanthropy.

The pupils were instructed to rank these in order of their choice. Each item was given a score of 2 if correctly placed in the ballot,

TABLE 68
INTERCORRELATIONS OF SERVICE TESTS *

	1	2	3	4	5
1. Free choice20	.17	.13	.20
2. Efficiency cooperation27	.32	.21
3. Money vote27	.12
4. Kits12
5. Envelopes					
Average inter-r201

* From Hartshorne, H., May, M. A., and Maller, J. B., *Studies in Service and Self-Control* (1929). By permission of The Macmillan Company, publishers.

a score of 1 if misplaced by one rank, and a score of 0 if misplaced two or more ranks.

Inhibition

A set of tests was devised by the Character Education Inquiry to measure inhibition. Only three of the devices will be described. The reader is referred to Hartshorne, May, and Maller's *Studies in Service and Self-Control* for complete descriptions.

In the "Story Inhibition Test" each child holds a copy of a very exciting story which is read aloud by the teacher. At the climax the story continues on the next page. The children are given the alternative of counting words on the page just read in order to make a score on a test, or opening the folders which are stuck together to find out the end of the story. A score of 1 is given if a pupil does not break open the seal. Two stories are used.

In the "Safe Manipulation Test" a toy combination safe was placed on each pupil's desk with instructions not to touch it until later when it was to be used, because if any pupil touched his immediately he would get an unfair advantage. Then a series of six paper-and-pencil tests was given. The examiner checked the dial of each safe between each test to see if it had been touched by the pupil. Scores were assigned according to the number of times the dial was in a different position on the six inspections.

The "Puzzle Manipulation Test" consisted of a box with various puzzles carefully placed on a puzzle-peg board so that if any article was disturbed it could be recognized. As in the pre-

TABLE 69

INTERCORRELATIONS OF INHIBITION TESTS *

	2	3	4	5	6
1. Stories277	.073	.320	.050	.210
2. Safes500	.256	.001	.384
3. Puzzles307	.177	.325
4. Pictures420	.240
5. Puzzles010
6. Candy					
Average inter-r	.237				

* From Hartshorne, H., May, M. A., and Maller, J. B., *Studies in Service and Self-Control* (1929). By permission of The Macmillan Company, publishers.

328 Diagnosing Personality and Conduct

vious test a box containing the puzzles was placed on each desk, and directions were given not to molest it while certain paper-and-pencil tests were tried. Later the boxes were removed to another room where they were inspected and scored.

The Story Inhibition Test has a total reliability of .65, and one manipulation test correlated with another about .50.

Measures of Caution

"Caution" has been studied by Brown (37) and Manson (38) in connection with intelligence tests only. It has been observed in the taking of objective examinations that there is a twilight area between the questions which are answered with every certainty that they are correct and those questions which are omitted altogether because the person taking the test is sure that he does not know their answers. The response made in this twilight zone differs according to the way the test is scored. If the test is scored according to the number of right responses, there need be no tendency to show caution, for every correct answer counts, whether obtained by knowledge or chance. But if the test is scored by the formula "right minus wrong," where an individual is penalized by making a wrong answer, the rashness-caution factor is apt to be exhibited. Some persons under these circumstances will attempt every item, confident that the vague associations which they are able to make in connection with the item will lead them to give the correct answer. Others exhibit extreme "caution" and refuse to answer any items except those of which they are certain.

Brown (36, 37) measured "caution" in those tests on the Thorndike "Intelligence Test for High School Graduates" in which the subject is warned that a wrong answer will count off from his score because the number of zero and minus scores will be counted—that is, the number of wrong answers. The argument here is that (37, p. 46) "the person in whom the caution factor is operating at its maximum would most probably refuse to write down any answer of which he was not absolutely sure. Lack of the knowledge that he is right, even when he has a high degree of confidence in his answer, would tend to make the extremely cautious individual count the item altogether out, while a less cautious person would hazard a guess in many cases." In the

Binet vocabulary test and the Trabue completion test, the caution index was expressed as the ratio of number of items wrong to number of items tried.

Brown found that there was a considerable correlation between his caution index and intelligence and also some with scholarship (.27 and .15 respectively), but that the correlation with conduct and the time taken in the test is so small as to be negligible.

As will be seen in the discussion under the measurement of studiousness, it is difficult to make a satisfactory study of personal characteristics by means of test results because of the statistical unreliability which follows when test scores are combined by taking differences or quotients. One suspects that the measures of caution employed by Brown also have a large element of chance which tends to reduce their reliability.

Speed of Decision

Another possible contact lending itself to study by testing is "speed of decision." It has been noticed that persons differ in their ability to make a choice or render a decision. Some seem to be so confused by the relative advantages and disadvantages of the issue at question that a decision is difficult. Just when the balance seems to be surging in one direction, some other point will spring into prominence to counterbalance it and block the issue. Other persons, on the other hand, seemingly have little difficulty in weighing the issues and estimating their relative importance, thus speedily reaching a decision.

In addition to the decision test in the Downey "Will-Temperament Test" to be reported upon, studies in decision have been made by Bridges (39), Gibson (41), Filter (40), and Trow (42). Of these studies the one that enlightens us most concerning the value of tests of speed of decision is that by Trow. He used eight tests as follows:

1. *Line discrimination.* Ten cards were prepared, each of which had a vertical line 100 mm. in height resting on a horizontal line which varied in length in the different cards from 95 to 105 mm. The subject was instructed to compare the length of the two sections, and the time after exposure of the card for such decision was recorded.

330 Diagnosing Personality and Conduct

2. *Weight discrimination.* Eight standard weights varying from 84 to 112 grams were used, a 100-gram weight serving as a standard of comparison. Weights were presented in pairs, one of each pair being the 100-gram weight, and the subject was instructed to state which was the heavier, the time of making the decision being recorded.

3. *Spelling.* Twenty difficult words in spelling were presented, eleven correctly spelled and nine with slight misspellings. The subject was instructed to indicate by marking with an R or a W whether the word was correctly or incorrectly spelled.

4. *Ethical judgment.* A blank with twenty questions such as, "Is capital punishment ever right?" was used. The subjects were instructed to answer each question by writing Y or N, standing for *yes* or *no*, before each question.

5. *Belief.* The Summer list of beliefs was used. A sample question is, "1. Is the world becoming better?"

6. *Rating.* Names of subjects in the experiment who were members of a psychology class were typed on cards. Each subject was instructed to arrange these cards in order, placing at one end the individual in the group judged to be the most self-confident and at the other end the individual judged to be the least self-confident.

7. *Speed of decision.* This test is taken from the Downey "Will-Temperament Tests."

8. *Finality of judgment.* This is also one of the Downey will-temperament series.

Twenty-seven subjects were given this battery of eight tests, each being a measure of the speed of decision. Reliability coefficients are not given. Intercorrelations were as follows:

TABLE 70
CORRELATIONS BETWEEN THE DIFFERENT MEASURES FOR SPEED OF DECISION

(from Trow, 42, p. 541)

	2	3	4	5	6	7	8
1. Line discrimination55	.44	-.08	.21	.29	.20	.29
2. Weight discrimination29	.13	.42	.40	.14	.08
3. Critical judgments17	.21	.29	.40	.14
4. Belief52	.29	.27	-.25
5. Rating41	.44	.33
6. Speed of decision44	-.04
7. Finality of judgment23

From these figures Trow concludes (42, pp. 541, 542), "The lack of what may well be called *trait consistency* is the conclusion that is forced upon us by this study of the speed of decision. . . . It seems clear from the above data that the persons who are quick to decide in some cases are slow in others, and *vice versa*. The man who is quick to decide to buy a house might be slow to decide which suit to put on in the morning. The man who is quick to decide against being a party to a shady deal might be slow to decide whether to fire a young employee caught in some petty dishonesty." The measure of time taken to make one decision is no dependable indication of the time necessary to make another decision.

Filter (40) in a similar study comes to much the same conclusions. When he used performance tests, his intercorrelations were very low. When he used a paper-and-pencil test, his correlations are considerably higher. Of this Filter says (40, p. 314):

"Correlations between test results are positive and fairly high, indicating that individuals who are quick in decisions of one kind tend to be quick also in decisions of other kinds. This does not constitute a demonstration of high degree of constancy of speed of decision, however, as later qualifications show. There is little likelihood that any one or two tests can be developed to measure the trait adequately."

Aggressiveness

Another quality which has attracted the interest of psychologists is aggressiveness, a trait supposedly of particular importance in industry. Since salesmen, executives, and the like apparently achieve success in part by aggressive behavior, a method for appraising this trait more accurately should be helpful in personnel work in industry and in vocational guidance.

An interesting experimental attack on the problem of the measurement of aggressiveness was made by Moore and Gilliland (44). They define the quality in which they are interested as follows: "He is more likely to be vigorous, positive, and masterful than the man lacking in this trait; and he is less likely to shrink from notice, to avoid argument, to display a lack of 'nerve'."

1. *Test of eye control.* Of the several tests which the experimenters used one was a test of eye control. It is commonly believed that ability to maintain a steady gaze is one characteristic

332 Diagnosing Personality and Conduct

of the man of power, and that a shifty eye is a sign of personal weakness. The test consisted of performing a somewhat difficult series of mental additions while constantly returning the fixed gaze of the instructor who sat facing him. The subject was emphatically instructed that under no circumstances should he let his gaze wander from that of the man facing him, as all movements of the eyes were to affect his score seriously. The number of eye movements is recorded as the test score. Thirteen men rated high in aggressiveness were found to make only 41 eye movements as compared with 72 such movements for the thirteen least aggressive.

2. *Fear distraction tests.* Subjects were practised in the addition tests until the practice effect was apparently negligible. The time in the eye control test previously mentioned when five addition series were used was then compared with the time for five addition series under normal conditions. For the aggressive subjects the staring caused an average delay of .4 of a second; for the unaggressive subjects it caused an average delay of 3.2 seconds.

A second series of distraction tests was given using as before the five series of addition tests as material and threatening the subject with the expectation of an electric shock which was to come during or at the end of each of the five series of additions. Although told that the shock might be from 75 to 220 volts, actually it was never more than 75 volts and was always given after a series was completed so that it did not actually interfere. The average shock-delay of the least aggressive subjects was 6.1 seconds as compared with an average of 2.2 seconds for the aggressive group.

In a third series a dead snake, suitably coiled and pinned to a cork board, was placed about ten inches in front of the face of the subject while he was adding. This stimulus caused an average delay of 8.2 seconds for the least aggressive subjects as compared with a delay of 4.6 for the aggressive ones.

Word association tests were also used, but these will not be described here.

Since variability figures are not given, there is no way of checking up on the significance of these differences. There is apparently correlation between the different tests used, but the size of the correlation is not stated.

Gilliland (42) later revised the test and partially standardized

it, omitting distraction by electric shock and by the snake and adding a test comparing writing at the normal rate and writing as fast as possible. The increase in speed of writing is used as a measure of aggressiveness. Correlations between the total measure and aggressiveness and other factors are: intelligence $+ .02$, scholarship $+ .02$, private lessons in speech $+ .34$, selling ability (newspaper advertising) $+ .26$. The correlation with grades in private lessons in speech were found because aggressiveness is commonly spoken of as a characteristic of the forceful and effective public speaker. Intercorrelations between the different tests of aggressiveness are not given.

Further work on measures of aggressiveness is needed. They seem to have some diagnostic power but it is not known to what degree the tests depend on the special conditions under which they are given.

Studiosness or Effort in School

In 1920 Franzen (49) devised a measure which he called the "accomplishment quotient." Almost simultaneously Monroe and Buckingham, in connection with the Illinois Intelligence Examination, proposed a similar measure which they called the "achievement quotient." Later the name was changed to "accomplishment ratio" so that it would not be confused with other quotients such as the intelligence quotient, in which the denominator of the fraction is chronological age. The accomplishment ratio is the ratio of educational achievement to mental development. If a child's educational achievement is greater than his mental development compared with other children of the same age, it is a symptom that he has worked harder than the average child at his studies. There seems to be no way that a child could surpass in educational achievement other children similarly placed and of the same mental development except by greater effort or industry. On the other hand, if the educational achievement is relatively lower than the mental development, a number of factors may be the cause. The child may have neglected his school work or been deprived of educational opportunity, or been affected by other matters interfering with his school progress.

In finding the accomplishment ratio, educational achievement is usually measured by the educational age. Educational age is

found from tables showing the average score on a test made by pupils of different ages. Usually several tests representing a variety of school subjects such as reading, arithmetic, and the like are combined in an equitable way before the educational age is determined. Mental age is similarly formed from the scores on an intelligence test by reading from a table the average scores made by pupils of different ages. The formula for the AR is $\frac{EA}{MA}$.

A similar constant suggested by Franzen (48) is found by subtracting mental age from educational age, but this difference has not found such wide usage as the ratio.

After adolescence is reached and the rate of learning and mental growth slows down, the units *mental age* and *educational age* become meaningless, and it is not possible to determine them except by artificial extensions of the tables of age norms. Moreover age loses its potency as a base in determining achievement in the secondary school subjects. A unit to serve in such situations has been proposed by Feingold (47) and Symonds (52). This "index of studiousness" is described by Symonds* as follows (54, pp. 521-524):

"If it were possible to eliminate from the achievement test scores the effect of intelligence test scores, the result, according to the argument and within the limits of test reliability, would be the effect of environment which we may call *effort* or *studiousness*. Looking at the matter in another way, intelligence test scores give a measure of 'average' ability, and predict what may be expected in the way of achievement. If more or less than the expected achievement results, the difference may be ascribed to greater or less effort or greater or less studiousness.

"The elimination of intelligence may be accomplished in several ways. One method is to turn both test scores and intelligence test scores into ranks and to obtain the difference between the ranks. This difference between ranks is the measure of studiousness desired.

"A second method for obtaining this measure of studiousness is to obtain standard deviations and the difference between standard deviations.

"The results using the two methods are in agreement in the main. The first method using the difference between rank is the easier method and is preferred by some for that reason. Both methods show a regression effect. That is, it is extremely difficult,

*From Symonds, P. M., *Measurement in Secondary Education*. By permission of The Macmillan Company, publishers.

statistically, for a pupil with high intelligence to make a plus studiousness index, and vice versa. That is, indeed, impossible in the method using ranks. The pupil having a rank of 1 in intelligence can have no lower rank than 1 in achievement, and hence the largest difference for this pupil is zero. Likewise the lowest possible achievement difference for the pupil lowest in intelligence is zero. This fault does not occur in the method using S. D. differences and the S. D. difference is the one recommended.

"This measure of studiousness suffers from a fault common to most measures of conduct. It differentiates between the members of a single class but does not permit comparison between members of different classes. Only when one or more individuals are common members of two groups can a comparison between groups be made. Assumptions about equal average studiousness of different groups are fallacious, as so much depends on the motivation and the social setting. What is needed in this connection are standards on standardized tests which will permit the comparison between groups. Such standards have been determined for nine of the best achievement tests that measure high school subjects."

The advantage of having a measure of effort or studiousness is apparent. Probably no factor has more influence in determining school achievement than native ability, but in the long run and within limits what is variously called *industry* or *application* or *effort* or *studiousness* can add to or subtract from what might be expected from the average use of ability.

The use of ratios or differences in studying the relative standing on two tests has been studied by Chapman and Kelley. Both find the method open to grave dangers. In the first place they point out that the order of unreliability of the differences between two test scores is considerably larger than the unreliability of the tests themselves. Chapman (46) has devised a formula which gives a reliability coefficient of the difference or quotient of two tests.

$$r_{a_1 a_2} = \frac{\frac{r_{11} + r_{22}}{2} - r_{12}}{1 - r_{11}}$$

Where $r_{a_1 a_2}$ is the reliability coefficient difference between an intelligence and achievement test

r_{11} is the reliability of the intelligence test

r_{22} is the reliability of the subject-matter test and r_{12} is the

336 Diagnosing Personality and Conduct

correlation between the intelligence test and the subject-matter test.

Using the formula on representative data, Chapman finds the reliability coefficient of the difference between an intelligence test and an achievement test to be very close to zero.

Kelley (50) independently attacked the same problem and has given us a formula for the reliability of the difference between the scores that an individual makes on two separate tests. First the two scores must be transmuted into standard scores and then the reliability of the difference may be determined by the formula

$$o_d = \sqrt{2 - r_{1I} - r_{2I}}$$

Kelley in addition to pointing out the unreliability of the differences between scores on intelligence and achievement tests has also discovered that an intelligence test and a general achievement test measure so nearly the same thing that differences between them are all but meaningless. Kelley says (51, pp. 21, 22),

"On the average, in the neighborhood of .90 of the capacity measured by an all-round battery score—reading, arithmetic, science, history, etc.—and of the capacity measured by a general intelligence test is one and the same. If a comprehensive educational achievement test and a general intelligence test each give 'fairly reliable' total scores, each would need to be more than ten times as long to yield equally reliable measures of difference between the educational achievement and the intelligence scores. This is true not only because 90 per cent of the tests measure a common function, but also because the chance factors entering into this 90 per cent of each test tend to obscure whatever real difference is being measured by the 10 per cent. This means that a scant one tenth of the tests are involved in the measure of difference, and, practically, that judgments of individual differences between intelligence and achievement based upon the commonly available tests are quite unsound, being of an order of accuracy not of the total scores of the tests, but of total scores of tests less than one tenth as long. The possibility of making sound judgments of this sort by utilizing much more refined measures lies before us."

Thus as the matter stands the theory underlying the accomplishment ratio or the studiousness index is attractive, but the actual intelligence and achievement tests which we now use give measures which do not possess the differentiation necessary to make the use of these derived scores satisfactory.

The Downey Will-Temperament Tests

The chapter or section of a chapter in the history of psychology which will deal with the Downey "Will-Temperament Tests" should be at the same time pathetic and amusing—pathetic because of the way in which gullible testers rushed in to use an instrument for which much had been promised, and amusing because of the politeness with which it was treated by scientific workers who had evidence of its worthlessness. Spurred on by the remarkable success attending the development and use of intelligence tests in the United States Army, psychologists after the World War began confidently casting about for techniques with which to measure other areas of personality. When Downey in 1919 announced a test of will-temperament, it was hailed enthusiastically and given willing trial by a large number of investigators. The name of the test itself was intriguing. Many psychologists accepting at face value Miss Downey's claims, made collections of "profiles" and attempted to interpret them. Others, more skeptical and experimentally minded, put the will-temperament tests to the usual tests of reliability and validity. No other test described in this book has had such widespread investigation, carried out on a very high level of painstaking accuracy. But from the very start the returns were discouraging, and they have consistently told the same story. Something is the matter with our scientific educators when they will remain so courteous and so hopeful in the face of evidence so convincing. Freeman, for instance, concludes after reviewing the tests and the work which has been done upon them (69, p. 105), "It is evident from the facts which have been presented that the Downey Will-Temperament Test, which is the most carefully standardized and most highly elaborated personality test which has yet been devised, is unsuitable for widespread routine application in the school. It is still in the experimental stage." Must we always remain in a state of hope and expectancy concerning measuring instruments which yield poor results on preliminary trials? Even such an intrepid investigator as May timidly concludes (78, p. 39), "The results of all those attempts at validation by the rating method are uniformly ambiguous. Nearly all the correlations are low. Does this mean that the tests are not valid, or the ratings unreliable, or that this is not the proper method of testing the

338 Diagnosing Personality and Conduct

tests? An examination of the data reveals the fact that these ratings are about as reliable as ratings usually run. Other kinds of tests have been validated by this method. We venture the assertion that after ample allowance has been made for errors in ratings the correlations will still be low." Are we never to be able to call a spade a spade?

In 1919 a bulletin from the University of Wyoming appeared entitled "The Will-Profile: A Tentative Scale for Measurement of Volitional Pattern" by June E. Downey (59). Herein was described a series of tests, based largely on controlled handwriting, purporting to measure qualities of will and temperament. Miss Downey wrote her doctor's dissertation in 1908 on *Control Processes in Modified Handwriting* in which she described experiments in handwriting changes under various conditions, such as when the writer is blindfolded or attempting to do something else at the same time, or when there is a time-lapse between the directions for writing and the actual writing. In previous work she had practised *muscle-reading* and had become remarkably skilful in responding to the involuntary cues provided by psychologically unsophisticated guides who concentrated on concealed objects. These studies led her to recognize differences in "motor impulsion," "resistance to opposition," "motor inhibition," "coordination," and "perseveration." She says,

"I discovered it was a simple matter to select exceptionally good guides for demonstrations in muscle-reading by a preliminary trial of writing under distraction of attention. The 'good guides' [those whom Miss Downey could easily follow by muscle-reading to the concealed object] under such distraction produce an *enlarged and rapid hand*. Extensive observation has thoroughly convinced me that this magnified and semi-automatic writing produced by the simple device of writing when the attention is partially distracted by a concurrent process distinguishes the impetuous individual for whom the motor discharge takes place easily and readily. Writing, decreased in size or greatly retarded in speed, appears under the same circumstances for those individuals who 'hold on to themselves painfully,' who do not yield 'readily to automatism,' who respond with increased effort of attention to the demand to handle a double process. The latter are the inhibited, obstructed, pondering type of individual."

Out of such observations were born the will-temperament tests. The original Downey Will-Temperament Test published in

1919 was an individual test. In 1921 it was published by the World Book Company, and in 1922 the World Book Company also published a group Will-Temperament Test. At about the same time (1920) these tests were used in the personnel investigations under way at the Bureau of Personnel Research of the Carnegie Institute of Technology in Pittsburgh and a form known as the "Carnegie Adaptation of the Will-Temperament Tests" was there developed by Ream. In 1925 a "Non-Verbal Will-Temperament Test" was developed by Downey and Uhrbrock.

In the original test there were twelve tests divided into three groups. The first group, composed of four tests, is designed to measure the "fluidity" or *speed of response* and to determine whether an individual is of the speedy, "hair-trigger," fluid type or the cautious, deliberate type. The second group of four tests measures *forcefulness* and *decisiveness of action*. Individuals are here thought of as varying from those who react in a forceful, decisive, determined way to those who are indecisive and easily led or controlled. The third group of four tests measure *carefulness* and *persistence of reaction*—the ability to attend to details and to continue a task for a long time.

Each of the tests will now be briefly described.

1. *Speed of movement.* In this first test the subject writes the words "United States of America" at his ordinary speed. As the test is given to individuals, the score is the time taken to perform the writing. In the group test there is a time-limit of twenty seconds, and the score is the number of letters written.

2. *Freedom from load.* The significance of this test would perhaps have been clearer if it had been called "freedom from inhibition." The theory underlying the test is that some people habitually write near their maximum speed, while others habitually write at a speed considerably below their maximum. The explanation given for this is that these latter persons are subject to "load" or "inhibition." For this test the subject is instructed to write "United States of America" as rapidly as possible and (in the group test) as many times as possible from the signal to begin to the signal to stop. The examiner is told to suggest speeding by voice and manner in the group test. This test is scored in the individual test by counting the seconds required for writing, or in the group test in the number of letters written

in the twenty seconds allowed, and then in both cases taking the ratio between speeded and normal writing.

3. *Flexibility.* In this test the subject is instructed to write "United States of America," changing the style of writing as much as possible "so that none of your friends would know it." This test is scored 0, 1, or 2 according to the degree to which the subject is able to change his style of writing as may be judged by comparing it with a scale of specimens. This test is supposed to measure either the dramatic or histrionic type of temperament or the exercise of ingenuity.

4. *Speed of decision.* In this test the subject is given twenty-two pairs of opposite traits (in the group test, thirty) and is asked to check the trait in each pair which characterizes himself. The score is the speed with which the judgments are made (either the time taken to check the whole list in the individual test, or the number checked in forty-five seconds in the group test). The purpose of this test is to determine the speed with which a person makes judgments of this kind.

In the next group of four tests we have:

5. *Motor impulsion.* The subject is requested to write his name in his usual manner, then with his eyes closed, then with his eyes open at the same time counting by threes, thus, 3, 6, 9, etc.; and finally to write his name while counting by twos. In the group test the order of events is writing the name (a) with eyes open, (b) with eyes closed, (c) with eyes fixed on a pencil held by the examiner, at the same time counting aloud the number of times the pencil is tapped on the table, and (d) writing the name and at the same time keeping track of the number of times the word *fly* is spoken by the examiner when he reads a list of words that rime with *fly*, such as *die*, *sigh*, and *lie*. This test is scored by reference to a table which takes into account both speed and size of writing. Miss Downey's previous researches showed that some individuals who scored higher on "motor impulsion" or "muscular tension" than average tend to speed up and enlarge the handwriting when working under distraction.

6. *Reaction to contradiction.* Early in the test the subject is asked to choose between two envelopes. Later the subject is asked to state which envelope he chose, whereupon the examiner contradicts him and denies that that was the envelope chosen. The subject is scored high if he persists in maintaining that he was

correct in stating which envelope he had previously chosen. He is scored low if he weakly gives in and admits that he was mistaken.

7. *Resistance to opposition.* In this test a small obstruction (such as a pasteboard box) is placed in front of the pen-point, exerting enough pressure so that to continue writing will require considerable effort. The subject is told to write his name with his eyes closed. The test is scored qualitatively by comparing the behavior "resisting opposition" with that described in a scale. Since this test does not lend itself to group testing, it is not included in the group scale.

8. *Finality of judgment.* The pairs of traits on which the subject earlier rated himself is again presented to the subject, and he is given the opportunity to make any changes he wishes to. The score is the time consumed in rechecking, the assumption being that the longer time consumed in rechecking, the less satisfied the subject is with the original rating.

The two tests "Reaction to Contradiction" and "Resistance to Opposition" do not lend themselves to group testing, and in the group form have been replaced by two other tests. One of them, called "Self-Confidence," consists of sixteen true-false items which concern a list of ten words read by the examiner earlier in the test. The items of which the subject is absolutely sure he is instructed to *underline*, and the score on this test is the number of statements so underlined. A second test is called "Non-Compliance." In this the subject is scored according to the number of changes he makes in his answers to a true-false test when told that eight of the sentences are true and eight are false. These are intended to serve approximately the same purpose as the two tests for which they are substituted in the individual test.

9. *Motor inhibition.* Here the subject is told to write "United States of America" as slowly as possible. The score is the time taken. In the group test the subject is instructed to move his pencil as slowly as possible along a dotted line, and the score is the number of units traced in a given time. This test requires a great deal of self-control, and many persons become very irritated at the task.

10. *Interest in detail.* The subject is instructed to copy some handwriting printed in the test booklet. This is done twice, once

342 Diagnosing Personality and Conduct

as rapidly as possible and again at a normal rate, but copying as exactly as possible. The score is a combination of the difference in speeds and the degree to which the model is approximated.

11. *Coördination of impulses.* Here the subject is instructed to write the words "United States of America" on a line a little over an inch long. He is told to write very rapidly and to take care not to run over the line. In the group test this test is scored by the number of letters omitted or which run over the line. In the individual test the degree to which the time approximates the time for normal writing is also taken into account.

12. *Volitional perseveration.* In the test for flexibility in which the subject attempts to disguise his handwriting he is given a certain amount of time for practice. In the individual test he is scored according to the length of time he elects to spend in practice. In the group test six minutes of practice is allowed altogether. During this time the examiner is copying the numbers 5, 10, 15, etc., on the blackboard at the elapse of so many seconds, and the subject is instructed to copy in his test booklet the last number written on the blackboard as he looks up from his practice, thereby giving a record of the time.

In summary, the will-temperament tests attempt to determine fundamental characteristics of reaction in an individual, using handwriting as the testing medium. Handwriting is convenient for the purposes of testing and is also a semi-automatic response which seems to exhibit personality differences when under the influence of various distractions or directions. If the tests actually do what is claimed for them, they should be very valuable instruments for diagnosing important features of personality and should find immediate usefulness in personnel work in industry, education, and medicine.

Downey instructs that a total score is not to be found for the test as a whole, inasmuch as each separate test measures a different quality, and a total score would accordingly be meaningless. Rather she suggests plotting each score in a "profile," so that an individual's relative position in each test may be seen.

The first item in which we are interested in the study of the value of the tests is the reliability of each separate test. Several studies of reliability of the Downey Will-Temperament Test have been made. Reliability is difficult to determine with total

Performance Tests

343

TABLE 71

RELIABILITY COEFFICIENTS ON THE GROUP TEST, DOWNEY WILL-TEMPERAMENT TEST

(from Downey and Uhrbrock, 66, pp. 30-35)

Test	REPETITION ONE DAY APART				Average
	149 normal college women	42 junior high school boys	37 high school girls	136 boys	
1. Speed of Movement,					
II-1	.81	.86	.33	.79	.70
II-2	.80	.83	.79	.89	.83
VI-1	.73	.82	.62	.50	.67
VI-2	.81	.76	.78	.81	.79
2. Freedom from Load,					
VI-1 and 2	.72	.39	.31		.44
VI-2 ÷ VI-1				.35	
3. Flexibility Test, VIII	.75	.67	.76	.31	.62
4. Speed of Decision, Test I	.63	.64	.64	.49	.60
5. Motor Impulsion Test, X-1	.72	.75	.61	.51	.65
(Total letter count)					
Test X-1 (av. millimeter measure)	.89	.89	.84	.76	.85
Test X-2, 3, 4 (av. letter count)	.79	.67	.72	.58	.69
Test X-2, 3, 4 (av. millimeter measure)	.80	.84	.84	.84	.83
Sum of ratios (performance in test X-2, 3, 4, ÷ Test X-1, for letter count and millimeter measure)				.23	
6. Self-Confidence Test XI	.51	.09	.57	.12	.32
7. Non-Compliance, Test XII	.31	.36	.50	.36	.38
8. Finality of Judgment, Test XIII	.46	.37	.26	.49	.40
9. Motor Inhibition, Test, III	.67	.44	.46	.66	.56
VII-3	.73	.76	.68	.63	.70
10. Interest in Detail, Test IX	.42	.66	.64	.41	.53
11. Coordination of Impulses, Test V	.47	.68	.56	.38	.52
(score in millimeters)					
12. Volitional Perseveration, Test VIII-2	.33	.40	.32	.34	.35
Average	.650	.626	.592	.522	.602

satisfaction on these tests. In the first place, it is unfair to judge reliability on the basis of repetition of many of the tests, for memory would so operate as to vitiate the results. In the case of the character rating or length of time of practice in disguising one's handwriting, memory would prevent the correlation of the two tests from being a true measure of reliability. On the other hand most of the tests cannot be split into halves. Downey and Uhrbrock (66) and later Uhrbrock (97) present correlations on the repetition of the test which are the best that we have to serve as reliability coefficients.

These correlations are highest where the test score is a measure of speed, where the skill tested is one well practised, where it is objectively measured, and where larger amounts of material make up the test, or where there are longer time limits. The facts show that the average reliability is greater for adults than for junior high school pupils, and that the correlations are higher when the interval between tests is a day than when it is a month, but it is not possible to generalize on these two points.

The reliability of the tests is not uniform, being considerably higher in some cases than in others. However, the reliability for a considerable number of the tests is sufficiently high so that the major criticism of the test cannot rest on this point.

The second critical check which may be applied concerns the intercorrelations of the tests one with another. These intercorrelations should be particularly significant within each of the three groups comprising the test battery, for the four tests in each of these groups are supposed to measure somewhat the same thing. The most comprehensive report of intercorrelations (Uhrbrock) contains tables which cannot be repeated here in detail. One will be given to indicate the nature of his results.

Uhrbrock, beside trying out the Downey Will-Temperament Tests, tried out many other tests which are called by the same name as the Downey tests. For instance, besides the test of speed of handwriting used by Downey, he tried out a variety of other tests of speed of movement such as tapping with right hand, tapping with right foot, tapping with left foot, reading simple prose rapidly, color-naming, etc. For each of these he gives intercorrelations, correlations with the Downey test, and correlations with a composite of all tests used. Similar work was done with

TABLE 72
INTERCORRELATIONS OF TESTS OF SPEED OF MOVEMENT
(from Uhrbrock)

	Indiv.	Verbal II-1	Verbal II-2	Verbal VI-1	Verbal VI-2	Non-verbal II-A	Non-verbal II-B	Composite of Carnegie fifteen tests I-10 movement of speed of
Verbal II-1	.37							+20
Verbal II-2	.21	.51						+49
Verbal VI-1	.26	.34	.37					+52
Verbal VI-2	.25	.30	.30	.46				+70
Non-Verbal II-A	.07	.08	-.21	-.07	-.16			-.13
Non-Verbal II-B	-.03	.11	.01	.06	-.01	.18		+11
Carnegie I-10	.49	.39	.49	.49	.55	.19	.18	+38
Carnegie 4	.39	.29	.57	.58	.58	-.04	-.02	+57

346 Diagnosing Personality and Conduct

other functions. In the following table average correlations are derived from correlations given by Uhrbrock:

TABLE 73
AVERAGE INTERCORRELATION OF THE DOWNEY WILL-TEMPERAMENT TESTS
(from Uhrbrock)

	<i>No. of correlations</i>	<i>Average correlation</i>
Speed of movement		
9 Downey tests	36	+.26
12 miscellaneous tests	66	+.21
9 Downey tests with 12 miscellaneous tests	96	+.08
9 Downey tests with composite	8	+.36
12 miscellaneous tests with composite	12	+.48
Motor inhibition		
5 Downey tests	10	+.34
3 miscellaneous tests	6	+.48
4 Downey tests with 3 miscellaneous tests	12	+.20
4 Downey tests with composite	5	+.50
3 miscellaneous tests with composite	3	+.56
Speed of decision		
5 Downey tests	10	+.34
2 miscellaneous tests	1	+.51
5 Downey tests with 2 miscellaneous tests	10	+.25
5 Downey tests with composite	5	+.59
2 miscellaneous tests with composite	2	+.60
Freedom from load, 4 Downey tests	6	.08
Flexibility, 3 Downey tests	3	.11
Motor impulsion, 4 Downey tests	6	.25
Self-confidence, 4 Downey tests	6	-.02
Non-compliance, 3 Downey tests	3	.05
Finality of judgment, 3 Downey tests	3	.19
Interest in detail, 3 Downey tests	3	.37
Coördination of impulses, 4 Downey tests	6	.26
Volitional perseveration, 4 Downey tests	6	.08

Ruch (88, 89) in an earlier study obtained the intercorrelations within each of the three groups of tests. The average intercorrelation for the four tests designed to measure the "hair-trigger" type of response was $-.005$; for the four tests supposed to indicate the wilful, aggressive type, $+.04$; and for the tests which purport to measure accuracy and tenacity, $+.06$.

If one may summarize on the basis of these averages, the intercorrelations are relatively low, so low indeed, that even in the Downey series one test cannot be said to measure the same thing as another test bearing the same name. The correlations with the composites indicate that the separate tests measure the true

function (if there be one) of which each test is a representative with correlations ranging from .40 to .60.

One must conclude in the first place that a serious error has been made in labeling the tests in the Downey series with general names such as "speed of decision" or "motor inhibition." The names are exceedingly technical and are not those that are in common usage even among psychologists. Consequently, it is impossible for most persons either to interpret the results of a profile so as to gather its significance, or to comprehend the terms used so as to recognize them in the person whose profile it is. "Freedom from load" means little to most persons until they have tried the test itself and see exactly what function is being tested. But this obscurity in the names of the tests is not their most serious criticism. The fact that tests bearing the same name correlate so low with each other is evidence that each test is measuring something very specific and that it is not possible to make inferences from the test results as to the personality of the individual.

This criticism of itself is sufficient to discourage the use of the Downey Will-Temperament Tests in any practical situation and to cast discredit on the value of any results obtained from their use. If, then, interpretations from the test results go no further than the tests themselves, of what value can they be? We are not directly interested in a practical way in how fast one can write or how slowly one can write, or whether one is able to disguise his handwriting. Data on these points are valuable only insofar as they indicate similar trends in the personality under circumstances that bear resemblance to the test situation. But the results show that this consistency is absent, and that a person may show speed in judging the taste of food or the height of a building and at the same time be slow in ascribing qualities to himself.

A third method of measuring the validity of the Downey Will-Temperament Tests is to determine the correlation between the tests and ratings on the qualities measured by the tests. Several studies have been made in which estimates of the qualities were obtained on groups of subjects to whom the tests had been given. Several psychologists, including Dr. Downey herself, have pointed out that these comparisons are not strictly fair, for such reasons as the well-known inaccuracies in the rating method, the

348 Diagnosing Personality and Conduct

difficulty in interpreting the names by which the Downey tests are known, and the difficulty in perceiving these qualities in others. However, these experiments have been carried through in good faith, and certainly the Downey tests ought to be subject to the same checks that are used on other tests.

The correlations obtained between the Downey Will-Temperament Tests and ratings of the same traits are given in the following table:

TABLE 74

CORRELATIONS BETWEEN SCORES ON THE DOWNEY WILL-TEMPERAMENT TEST AND RATING FOR THE SAME TRAITS

Downey Will-Tempera- ment Test	Ruch ⁽⁶⁴⁾ faculty estimates on 20 students	Ruch ⁽⁶⁴⁾ student estimates on 20 students	Ruch and Del Manzo ⁽⁶⁵⁾ pooled estimates on 146 high school students
A	.02	.43	.37
B	-.09	.27	.02
C	.45	.17	.30
D	-.02	.53	-.29
E	.51	.50	.10
F	.23	.37	-.03
G	.35	.28	-.04
H	-.11	.07	.17
I	-.33	-.23	-.09
J	.15	-.07	.51
K	-.09	.05	.53
L	-.26	.05	.17
Average	.07	.20	.14

Downey Will-Tempera- ment Test	Meier ⁽⁶⁶⁾ estimates by three judges on 106 high school students	Hershcovitz ⁽⁷²⁾ inter-ratings of three groups of persons, 7, 8, and 9 persons in each group	Average
A	.21	-.03	.20
B	.07	.18	.09
C	.11	.12	.23
D	.19	-.05	.07
E	.10	.21	.28
F	.05	.10	.14
G	.14	.18	.18
H	.14	.35	.12
I	.24	.29	-.02
J	.21	.00	.16
K	.07	.28	.17
L	.03	-.14	-.03
Average	.13	.12	.13

These correlations with ratings are consistently low and indicate that there is little relationship between the traits as measured by the tests and as estimated by careful observers.

With the main lines of evidence in hand we need go no further in order to draw conclusions as to the value of the tests. However, before doing so we wish to refer to two other types of experimental check worth recording. One bit of evidence prepared by Downey (61), fourth among the experimental checks, was the ability to recognize the person whose profile is under scrutiny. A group of judges was given twelve profiles and a list of the names of the persons to whom these belonged. The task was to match the name with the profile (61, p. 286). "Correct identification of profiles ran from 0 to 5 out of 12, or from total failure to identify any profile (1 judge) to 41 per cent of successful identification (2 judges). The percentage of successes for the total of 144 judgments (12 judgments by each of 12 judges) was 22, where chance success would be less than 1 per cent."

These results must be considered unsatisfactory. At another time Downey (61) submitted profiles in groups of three with instructions to select the profile that best fitted a given person. In series A the profiles were similar, while in series B they represented contrasting types. The average success of judges in series A was 33.3, 44.4, and 51.3 per cent, while in series B they were 78.7, 71.9, and 58.6 per cent. Certainly this is not a high percentage of successes and would hardly testify to the practical usefulness of the tests. Downey explains the failure partly on the ground that not only were the trait names partly ambiguous and obscure, but that the underlying temperamental traits themselves are difficult to detect by surface observation.

The sixth and last test of the validity of the tests is the degree to which they differentiate persons already known to differ according to race, sex, social adjustment, etc. The tendency among experimenters to record every difference, however small and however statistically insignificant, warns one not to accept published statements as to these differences without a challenge.

Bryant (56) found a difference of 15.5 points between delinquents and non-delinquents. May (78) estimates the P. E. of this difference to be 2.5, making the difference significant. Downey says of this, "This scale has evidently hit upon two or three of the more striking differences between delinquents."

Wires, studying sixty-seven psychopathic patients, concludes optimistically, "There were no instances in which a score as indicated on the profile was noticeably opposed to the personality of the subject as shown by the history or analysis." Bryant describes her delinquents as wilful, aggressive, and possessing a greater tendency toward accuracy and tenacity than toward adaptability. Wires describes her group as "impulsive, poorly inhibited, with relatively high scores on impulsion and assurance, and lowered scores for resistance and inhibition."

McFadden and Dashiell (79) have studied white and negro high school and college students and have found differences in almost all of the will-temperament tests between Indians and whites. Freyd (70) discovered that certain of the tests showed reliable differences between socially inclined and mechanically inclined groups.

It is difficult to see how two forms of the same test could give such low correlations and yet at the same time be such potent instruments for diagnosing differences between groups. Perhaps one should accept the evidence at its face value until repetitions of the experiments verify or fail to verify the findings.

The tests have also been used to study the discrepancy between scholarship and intelligence. Poffenberger and Carpenter (81) studied two groups, one the "success" group with higher rating for school success but lower ratings for intelligence; the other "failure" group with lower ratings for school success and higher ratings for intelligence. Differences between the groups were noted on the will-temperament tests. Tests were tried in all possible pairs, from which thirty-one pairs were selected as possessing differentiating ability. When combined they gave an average of plus scores of 9.4 to the success group and 3.5 to the failure group. There was also an average of minus scores of 3.6 for the success group and 9.9 for the failure group. To the present writer this seems like a grand capitalization of chance differences. It would seem as though all the chance differences existing in this particular set of data were ferreted out by careful statistical treatment and then summated to present a picture of the total possible effect of chance differences. Studies of this type need to be repeated before we can know to what degree the differences are real.

Uhrbrock (97) finds the average correlation between a test

of intelligence and the will-temperament tests taken singly to be $+.08$. Ruch and Del Manzo (89) report correlations between the will-temperament tests and an average of the Terman "Group Test of Mental Ability" and the Morgan "Group Mental Test." The average of these is $+.17$. Meier (80) reports a correlation of $+.21$ between "total score" on the Downey test and on the Terman group test. Bryant (55) finds correlations of $.38$ between total will-temperament and Stanford-Binet IQ, and $.43$ between will-temperament and Stanford-Binet mental age. The evidence indicates that the Will-Temperament tests correlate slightly and positively with tests of intelligence.

Summary of Downey Will-Temperament Tests. A brief categorical summary of the findings should be of service in helping to make a final estimate of the significance and value of the Downey Will-Temperament Tests.

1. *Reliability.* The tests vary in reliability from about $.30$ to over $.80$. The average reliability is about $.60$. Some of the tests involving handwriting are highly reliable. Others that involve less habitual reactions and in which the directions are not perfectly standardized are less reliable.

2. *Intercorrelations.* The intercorrelations of tests are very near zero. Even the correlations between tests bearing the same name are very low. The correlations of separate tests in the Downey series with composites of several tests supposed to measure the same functions range from $.40$ to $.60$. This would indicate that the functions measured by the Downey tests are specific and vary appreciably with each change in the situation or the task. There is grave doubt whether there actually exist general contacts in the majority of persons that correspond to the test names assigned by Downey.

3. *Correlations with ratings.* The correlations between separate Downey tests and ratings for these same traits in the individuals tested are very low. This may indicate one of several things: either the traits themselves do not exist, or the tests are poor measures of the traits whose names they bear, or the qualities are difficult to observe and rate in others.

4. *Groups separated by race.* Social adjustments and the like have been reported to produce or go hand in hand with differences in the responses to the items. If these findings are substantiated, they will point to the value of the tests regardless of the low

intercorrelations and the low correlations with ratings. Some of the differences reported were mere chance differences and need be given no further attention. Other experimenters have demonstrated the differences to be statistically significant. But even in these cases we do not know how important the differences really are, and it may turn out in a repetition of the experiment that chance exaggerated certain of the differences.

Conclusion. The general enthusiasm that originally greeted the Downey Will-Temperament Tests has subsided. Their practical value in guidance, personnel selection, and the like has not been demonstrated. Experimental work shows them to be tests of very specific abilities rather than tests to diagnose the more general personal qualities the names of which they bear.

Summary

From a review of all of the skilful and ingenious methods for testing conduct directly that have been devised, the conclusion stands out above all others that conduct is very specific. When exactly the same test is repeated, the correlation is fairly high, perhaps around .70 or .80. But when the situation is changed ever so slightly, the correlation between the two similar tests drops, and long before the two situations seem different enough to be called by different names, the correlation has dropped close to zero. A battery of tests designed to test such a trait as persistence, or aggressiveness, or speed of decision gives results so varying and with so little consistency as to furnish little warrant for assuming the presence of such a trait.

These low intercorrelations also help to explain the low correlations of these tests with outside criteria. Since the tests are so specific as to fail to correlate with tests bearing the same name, naturally they could not be expected to correlate to any degree with other factors which are admittedly dissimilar.

The conclusions that one draws from these results are not very encouraging. There are four possible things that may be done with performance tests in the measurement of conduct.

1. They may be discarded as being so specific as to be useless for all practical purposes.

2. Tests may be devised that apply to the specific situation in which they will be used. Since tests are so very specific, test

situations must be set up which approximate as closely as possible the situations in office, industry, school, or institution where they will be used. If aggressiveness in selling is what is wanted, then the test situation must be one involving selling.

3. Since no one test measures a given quality adequately, a variety of tests representing a range of situations in which the trait occurs may be devised so that the composite will be a satisfactory measure of the trait in question. This is what May and Hartshorne have done with the tests of *deceit* as well as with their later tests of *helpfulness*, *inhibition*, and *persistence*. Thirty-two tests of deceit combined into one composite should give a fairly representative measure of honesty. But honesty thus conceived is an abstract, composite affair. It, in turn, has little relationship to the specific act in the specific occasion for cheating. To know a person's honesty in general is not of high value in knowing how he will react in a specific situation.

We might follow this reasoning out to a more general conclusion. The measurement of character is a relatively useless and theoretical interest. Character is so very general as to give little inkling of what to expect in a specific situation. In other words, to select with confidence an honest bank clerk, we should test a candidate for honesty in handling money as a teller behind the grill in a bank. To know his honesty in any other situation or even to know his honesty in general is to know little about his honesty in the particular situation.

However, if it should happen that the more desirable the qualities a person possesses, the more integrated his conduct becomes, then tests would become valuable in selecting individuals. It may be found that to find a man honest or persistent or assertive in any one situation is high presumption that he will be found persistent or assertive in other situations. But the converse is not true, and the man who is deceitful, who gives up easily or is timid in one situation, will not necessarily exhibit these same characteristics in other situations. The criminal is one who usually has very strongly set habits which are specific in nature, but the good man is likely to be one whose conduct shows a higher degree of consistency.

4. A fourth method of using these tests is to pick out one test for each of a number of different traits and weigh these tests in combination in order best to predict success in business or

354 Diagnosing Personality and Conduct

school. Since the tests have fair reliability but low intercorrelations among themselves, it is very probable that a wisely selected group of tests weighted by the regression equation technique could be found to predict differences in delinquency or other forms of adjustment with a considerable degree of success.

In all of these alternative methods of using performance tests certain practical problems of cost, difficulty of administration, difficulty in applying statistical techniques, and the like, arise which very definitely limit the use of these tests. It is often expensive to test in the practical situation. Certain ingenious devices must be applied which eat into time and money, as May and Hartshorne found out. Again, to give a well-rounded battery of tests is also expensive. Finally, one who plans to use the regression equation technique must count the cost beforehand.

Performance tests have a real and valuable place at the present time in experimental work. As used by Hartshorne and May they have revealed facts that were obtainable in no other way. But there must be considerable further development before tests of this type become a feasible tool in clinical work.

REFERENCES

Measures of Deceit—Trustworthiness, Etc.

1. ATHEARN, W. S., *Measurements and Standards in Religious Education* (George H. Doran Company, 1924), Vol. II.
2. BIRD, C., "An Improved Method of Detecting Cheating in Objective Examinations," *Journal of Educational Research*, 19: 341-348 (May, 1929).
3. ———, "The Detection of Cheating in Objective Examinations," *School and Society*, 25: 261-262 (1927).
4. CADY, V. M., *The Estimate of Juvenile Incurability*, California Bureau of Juvenile Research, *Journal of Delinquency Monographs*, No. 2 (1923).
5. CLARK, F. A., "Some Character Tests," *American Educational Digest*, 46: 225-226 (1927).
6. COWEN, P. A., "Professional Spirit," *School and Society*, 26: 108-109 (July 23, 1927).
7. CUSHING, H. M., and RUCH, G. M., "An Investigation of Character Traits in Delinquent Girls," *Journal of Applied Psychology*, 11: 1-7 (1927).
8. FRANZEN, R., "Measurement of Non-intellectual Aspects of Behavior," *Proceedings First Annual Conference of Educa-*

- tional Record and Guidance, *San Jose Teachers College Bulletin*, 1922.
9. GUNDLACH, R., "A Method for Determining Cheating in College Examinations," *School and Society*, 22: 215-216 (August 15, 1925).
 10. HARTSHORNE, H., and MAY, M. A., *Studies in Deceit* (The Macmillan Company, 1928).
 11. MAY, M. A., and HARTSHORNE, H., "First Steps towards a Scale for Measuring Attitudes," *Journal of Educational Psychology*, 17: 145-162 (1926).
 12. MILLER, G. F., "An Experimental Test of Intellectual Honesty," *School and Society*, 26: 852-854 (1927).
 13. MURDOCH, K., "A Study of Differences Found in Races in Intellect and in Morality," *School and Society*, 22: 628-632, 654-664 (1925).
 14. PERRY, D. E., *A Measurement of Trustworthiness*, unpublished master's essay, Department of Psychology, Columbia University, 1923.
 15. RAUBENHEIMER, A. S., "An Experimental Study of Some Behavior Tests of the Potentially Delinquent Boy," *Psychological Monographs*, Vol. XXXIV, 6, No. 159 (1925).
 16. TERMAN, M., *Genetic Studies of Genius*, Vol. I (Stanford University Press, 1925).
 17. VOELKER, P. F., "An Account of Certain Methods of Testing of Moral Reactions in Conduct," *Religious Education*, 16: 81-83 (1921).
 18. ———, *The Function of Ideals and Attitudes in Social Education; an Experimental Study*, Teachers College Contributions to Education, No. 112 (1921).
 19. WITTY, P. A., and LEHMAN, H. C., "The So-called 'General Character Test,'" *Psychological Record*, 34: 401-414 (1927).
 20. WOODROW, H., and BEMMELS, V., "Overstatement as a Test of General Character in Pre-School Children," *Journal of Psychology*, 18: 239-246 (1927).
 21. YEPSEN, L. N., "The Reliability of Self-Scored Measures," *School and Society*, 26: 657-660 (1927).

Suggestibility

22. AVELING, F., and HARGREAVES, H. L., "Suggestibility with and without Prestige in School Children," *British Journal of Psychology*, 12: 53-75 (1921).
23. BROWN, W., "Individual and Sex Differences in Suggestibility," *University of California Publications*, 2: 292-440 (1916).
24. MCGEOCH, J. A., "The Relationships between Three Tests of Imagination and Their Correlation with Intelligence," *Journal of Applied Psychology*, 8: 439-443 (1924).

356 Diagnosing Personality and Conduct

25. MCGEOCH, J. A., "The Relationship Between Suggestibility and Intelligence in Delinquents," *Psychological Clinic*, 16: 133-134 (1925).
26. OTIS, M., *A Study of Suggestibility of Children*, Archives of Psychology, Vol. XI, No. 70 (1924).
27. PRIDEAU, E., "Suggestion and Suggestibility," *British Journal of Psychology*, 10: 228-241 (1920).
28. TOWN, C. H., "An Experimental Study of the Suggestibility of 12 and 15-Year-Old Boys," *Psychological Clinic*, 10: 1-12 (1926).
29. WHIPPLE, G. M., *Manual of Mental and Physical Tests* (Warwick and York, 1914).

Persistence

30. BRONNER, A. F., *Comparative Study of the Intelligence of Delinquent Girls*, Teachers College Contributions to Education, No. 68 (1914).
31. CHAPMAN, J. C., "Persistence, Success, and Speed in a Mental Task," *Pedagogical Seminary*, 31: 276-284 (1924).
32. FERNALD, G. G., "An Achievement Capacity Test," *Journal of Educational Psychology*, 3: 331-336 (1912).
33. HARTSHORNE, H., MAY, M. A., and MALLER, J. B., *Studies in Service and Self-Control* (The Macmillan Company, 1929).
34. MORGAN, J. J. B., and HULL, H. L., "The Measurement of Persistence," *Journal of Applied Psychology*, 10: 180-187 (1926).
35. RONDELLI, A., "Sur Valore Psicometrico dei Givochi Infantili," *Arch. di Antrop. Crim.*, 48: 19-29 (1928).

Caution

36. BROWN, W. M., "A Study of the 'Caution' Factor and Its Importance in Intelligence Test Performance," *American Journal of Psychology*, 35: 368-386 (1924).
37. ———, *Character Traits as Factors in Intelligence Test Performance*, Archives of Psychology, Vol. XI, No. 65 (May, 1923).
38. MANSON, G. E., "Personality Differences in Intelligence Test Performance," *Journal of Applied Psychology*, 9: 230-255 (1925).

Speed of Decision

39. BRIDGES, J. W., *An Experimental Study of Decision Types and Their Mental Correlates*, Psychological Monographs, Vol. XVII, No. 72 (Aug., 1914).
40. FILTER, R. O., "An Experimental Study of Character Traits," *Journal of Applied Psychology*, 5: 297-317 (1921).

41. GIBSON, S. M., "A Decision Study of 150 Young Men and Women," *Journal of Applied Psychology*, 4: 364-374 (1920).
42. TROW, W. C., "Trait Consistency and Speed of Decision," *School and Society*, 21: 538-542 (May 2, 1925).

Aggressiveness

43. GILLILAND, A. R., "A Revision and Some Results with the Moore-Gilliland Aggressiveness Test," *Journal of Applied Psychology*, 10: 143-150 (1926).
44. MOORE, H. T., and GILLILAND, A. R., "The Measurement of Aggressiveness," *Journal of Applied Psychology*, 5: 97-118 (1921).
45. RIDDLE, E. M., *Aggressive Behavior in a Small Social Group*, Archives of Psychology, Vol. XII, No. 78 (1925).

Studiousness

46. CHAPMAN, J. C., "The Unreliability of the Difference between Intelligence and Educational Ratings," *Journal of Educational Psychology*, 14: 103-108 (Feb., 1923).
47. FEINGOLD, G. A., "The Measurement of Effort among High School Pupils," *Educational Administration and Supervision*, 10: 385-394 (1924).
48. FRANZEN, R., *The Accomplishment Ratio*, Teachers College Contributions to Education, No. 125 (1922).
49. ———, "The Accomplishment Quotient of School Marks in Terms of Individual Capacity," *Teachers College Record*, 21: 432-440 (Nov. 1920).
50. KELLEY, T. L., "A New Method for Determining the Significance of Differences in Intelligence and Achievement Scores," *Journal of Educational Psychology*, 14: 321-333 (Sept., 1923).
51. ———, *Interpretation of Educational Measurements* (World Book Company, 1927).
52. SYMONDS, P. M., "A Measure of Studiousness," *Pedagogical Seminary*, 32: 257-265 (1925).
53. ———, *Ability Standards for Standardized Achievement Tests in the High School* (Teachers College Bureau of Publications, 1927).
54. ———, *Measurement in Secondary Education* (The Macmillan Company, 1927).

Downey Will-Temperament Tests

55. BRYANT, E. K., "The 'Will-Profile' of Delinquent Boys," *Journal of Delinquency*, 6: 294-309 (1921).
56. ———, "Delinquents and Non-delinquents on the Will-Temperament Test," *Journal of Delinquency*, 8: 46-63 (1923).

358 Diagnosing Personality and Conduct

57. CLARK, W. W., "Supervised Conduct-Response of Delinquent Boys," *Journal of Delinquency*, 6: 387-401 (1921).
58. COLLINS, M., "Character and Temperament Tests," *British Journal of Psychology*, 16: 89-99 (1925).
59. DOWNEY, J. E., "The Will-Profile," *Department of Psychology Bulletin*, No. 3 (University of Wyoming, 1919).
60. ———, "Ratings for Intelligence and for Will-Temperament," *School and Society*, 12: 292-294 (1920).
61. ———, "Some Volitional Patterns Revealed by the Will-Profile," *Journal of Experimental Psychology*, 3: 281-301 (1920).
62. ———, "Testing the Will-Temperament Tests," *School and Society*, 16: 161-168 (1922).
63. ———, *The Will-Temperament and Its Testing* (World Book Company, 1923).
64. ———, "Jung's 'Psychological Types' and the Will-Temperament Patterns," *Journal of Abnormal and Social Psychology*, 18: 345-349 (1924).
65. ———, "Observations in the Validation of the Group Will-Temperament Test," *Journal of Educational Psychology*, 18: 592-600 (1927).
66. DOWNEY, J. E., and UHRBROCK, R. S., "Reliability of the Group Will-Temperament Tests," *Journal of Educational Psychology*, 18: 26-39 (1927).
67. FILTER, R. O., "An Experimental Study of Character Traits," *Journal of Applied Psychology*, 5: 297-317 (1921).
68. FLEMMING, C. W., *A Detailed Analysis of Achievement in the High School*, Teachers College Contributions to Education, No. 196 (1925).
69. FREEMAN, F. N., "Tests of Personality Traits," *The School Review*, 33: 95-106 (1925).
70. FREYD, M., "The Personalities of the Socially and Mechanically Inclined," *Psychological Monographs*, Vol. XXXIII, 4, No. 151 (1924).
71. GARTE, T. R., and BARNARD, M. A., "The Will-Temperament of Indians," *Journal of Applied Psychology*, 11: 512-518 (1927).
72. HERSKOWITZ, M. J., "A Test of the Downey Will-Temperament Test," *Journal of Applied Psychology*, 8: 75-88 (1924).
73. HULL, C. L., and LIMP, C. E., "The Differentiation of the Aptitude of an Individual by Means of Test Batteries," *Journal of Educational Psychology*, 16: 73-88 (1925).
74. HURLOCK, E. B., "The Suitability of the Downey Will-Temperament Test as a Test for Children," *Journal of Applied Psychology*, 10: 67-74 (1926).
75. KOLSTEAD, A., "The Downey Will-Temperament Test in the

- Normal School," *Journal of Educational Research*, 10: 332-334 (1924).
76. KORNHAUSER, A. W., "Results from Testing of a Group of College Freshmen with the Downey Group Will-Temperament Test," *Journal of Educational Psychology*, 18: 40-42 (1927).
 77. KUNTZ, L. A., "A Study of the Literature on the Downey Will-Temperament Test," *Catholic Educational Review*, 23: 478-485 (1925).
 78. MAY, M. A., "The Present Status of the Will-Temperament Tests," *Journal of Applied Psychology*, 9: 29-52 (1925).
 79. McFADDEN, J. F., and DASHIELL, J. F., "Racial Differences as Measured by the Downey Will-Temperament Test," *Journal of Applied Psychology*, 7: 30-53 (1923).
 80. MEIER, N. C., "A Study of the Downey Test by the Method of Estimates," *Journal of Educational Psychology*, 14: 385-395 (1923).
 81. NACCARATI, S., and GARRETT, H. E., "The Relation of Morphology to Temperament," *Journal of Abnormal and Social Psychology*, 19: 254-363 (1924).
 82. POFFENBERGER, A. T., and CARPENTER, F. L., "Character Traits in School Success," *Journal of Experimental Psychology*, 7: 67-74 (1924).
 83. REAM, M. J., "Group Will-Temperament Tests," *Journal of Educational Psychology*, 13: 7-16 (1922).
 84. ———, "Temperament in Harmonious Human Relationships," *Journal of Abnormal and Social Psychology*, 17: 58-61 (1922).
 85. ———, *Ability to Sell* (The Williams and Wilkins Company, 1924).
 86. REAVES, W. C., "Utilizing the Results of the Downey Will-Temperament Tests in School Administration," *School Review*, 33: 174-183 (1925).
 87. ROE, A. M., and BROWN, C. F., "Qualifications for Dentistry," *Personnel Journal*, 6: 176-181 (1927).
 88. RUCH, G. M., "A Preliminary Study of the Correlations between Estimates of the Volitional Traits and the Results of the Downey 'Will-Profile,'" *Journal of Applied Psychology*, 5: 159-162 (1921).
 89. RUCH, G. M., and DEL MANZO, M. C., "The Downey Will-Temperament Group Test: Analysis of Its Reliability and Validity," *Journal of Applied Psychology*, 7: 65-76 (1923).
 90. STODDARD, G. D., and RUCH, G. M., "Ratings of Downey Will-Temperament Traits," *Journal of Applied Psychology*, 10: 421-426 (1926).
 91. STONE, C. L., "Disparity between Intelligence and Scholarship," *Journal of Educational Psychology*, 13: 241-244 (1922).

360 Diagnosing Personality and Conduct

92. SUNNE, D., "Personality Tests: White and Negro Adolescents," *Journal of Applied Psychology*, 9: 256-280 (1925).
93. THOMPSON, R. S., "A Study of the Validity of the Downey Will-Temperament Test," *Journal of Educational Psychology*, 19: 622-628 (1928).
94. TRAXLER, A. E., "The Will-Temperament of Upper-Grade and High School Pupils," *School Review*, 33: 264-273 (1925).
95. TROW, W. C., "Trait Consistency and Speed of Decision," *School and Society*, 21: 538-542 (1925).
96. UHRBROCK, R. S., and DOWNEY, J. E., "A Non-verbal Will-Temperament Test," *Journal of Applied Psychology*, 11: 95-105 (1927).
97. UHRBROCK, R. S., *An Analysis of the Downey Will-Temperament Tests*, Teachers College Contributions to Education, No. 296 (1928).
98. WIRES, E., "The Downey Will-Temperament Profile in Personality Studies of Juvenile Delinquents," *Journal of Abnormal and Social Psychology*, 20: 416-440 (1926).

Chapter X

THE FREE ASSOCIATION METHOD

THE free association method stands apart as unique among experimental techniques and as one of the most potent tools for the diagnosis of conduct. In essence the method consists of presenting the subject with a list of words, to each of which he is requested to respond by saying the first word which occurs to him. This makes a true measure of conduct because the response is perfectly free, except as it is required that the *first* word which is aroused by the stimulus word shall be given.

Sir Francis Galton, pioneer in so many matters psychological, was the first to make use of this method in a systematic way (1879). Wundt soon after experimented with the method in his laboratory. Beginning with Kräpelin in 1896, continuing through Aschaffenburg and culminating with the work of Jung and Riklin in 1906 came the noteworthy applications of the association method to the diagnosis of "complexes" or centers of emotional irritation. Another line of development was started by Münsterberg in 1889 through his use of the association method in the detection of guilt and dissimulation. In 1910 Kent and Rosanoff published their study of the use of the free association method in the detection of insanity. Since the method must be used individually and partly also because the exploitation of intelligence tests has engrossed psychological investigators, interest in it has recently somewhat died down, but in view of the potential fertility of the method, it is due for a revival.

As is so often the case, some of the sagest observations upon the method were made by the earliest workers. Galton (16, 17) observed, for instance, that free association is never entirely free. To a given word there is not an equal probability that any one of the many thousand other words in the English language will be given. Indeed, Galton discovered that for any individual the range of responses to a given word is decidedly small, being only

362 Diagnosing Personality and Conduct

a few words at the most. He also found, in his own case, that many of the associations were of long standing, having been formed in boyhood or early youth. Out of the myriad associations in which each word has appeared, the few to occupy an abiding place in the memory must receive many repetitions or have *continued living interest*. These early observations have been repeatedly confirmed in the later investigations. To quote one, Miss Fürst (34, ch. XI), a pupil of Jung, studied the degree to which associations tend to become constellated in families. Of this work Jung says (30, p. 245):

"One might indeed think that in the experiment, where full scope is given to chance, individuality would become a factor of the utmost importance, and that therefore we might expect a very great diversity and lawlessness of associations. But as we see the opposite is the case. Thus the daughter lives constantly in the same circle of ideas as her mother, not only in her thought but in her form of expression, indeed, she even uses the same words.

TABLE 75
KENT-ROSANOFF LIST OF FREE-ASSOCIATION WORDS
(37, pp. 37-38)

1. table	26. wish	51. stem	76. bitter
2. dark	27. river	52. lamp	77. hammer
3. music	28. white	53. dream	78. thirsty
4. sickness	29. beautiful	54. yellow	79. city
5. man	30. window	55. bread	80. square
6. deep	31. rough	56. justice	81. butter
7. soft	32. citizen	57. boy	82. doctor
8. eating	33. foot	58. light	83. loud
9. mountain	34. spider	59. health	84. thief
10. house	35. needle	60. bible	85. lion
11. black	36. red	61. memory	86. joy
12. mutton	37. sleep	62. sheep	87. bed
13. comfort	38. anger	63. bath	88. heavy
14. hand	39. carpet	64. cottage	89. tobacco
15. short	40. girl	65. swift	90. baby
16. fruit	41. high	66. blue	91. moon
17. butterfly	42. working	67. hungry	92. scissors
18. smooth	43. sour	68. priest	93. quiet
19. command	44. earth	69. ocean	94. green
20. chair	45. trouble	70. head	95. salt
21. sweet	46. soldier	71. stove	96. street
22. whistle	47. cabbage	72. long	97. king
23. woman	48. hard	73. religion	98. cheese
24. cold	49. eagle	74. whiskey	99. blossom
25. slow	50. stomach	75. child	100. afraid

TABLE 76

JUNG LIST (MODIFIED BY EDER) OF FREE-ASSOCIATION WORDS
(34, p. vii)

1. head	26. blue	51. frog	76. wait
2. green	27. lamp	52. try	77. cow
3. water	28. carry	53. hunger	78. name
4. sing	29. bread	54. white	79. luck
5. dead	30. rich	55. child	80. say
6. long	31. tree	56. speak	81. table
7. ship	32. jump	57. pencil	82. naughty
8. make	33. pity	58. sad	83. brother
9. woman	34. yellow	59. plum	84. afraid
10. friendly	35. street	60. marry	85. love
11. bake	36. bury	61. home	86. chair
12. ask	37. salt	62. nasty	87. worry
13. cold	38. new	63. glass	88. kiss
14. stalk	39. habit	64. flight	89. bride
15. dance	40. pray	65. wool	90. clean
16. village	41. money	66. big	91. bag
17. pond	42. silly	67. carrot	92. choice
18. sick	43. book	68. give	93. bed
19. pride	44. despise	69. doctor	94. pleased
20. bring	45. finger	70. frosty	95. happy
21. ink	46. jolly	71. flower	96. shut
22. angry	47. bird	72. beat	97. wound
23. needle	48. walk	73. box	98. evil
24. swim	49. paper	74. old	99. door
25. go	50. wicked	75. family	100. insult

What seems more flighty, more inconstant, and more lawless than a fancy, a rapidly passing thought? It is not, however, lawless, and not free, but closely determined within the limits of the milieu."

In giving the free association test, 100 words are generally employed. With fewer than 100 the results are not believed to be reliable. Some have stressed the point that only with a long list of words are the mode of reaction mechanized and the inhibitions reduced. On the other hand, since fatigue sets in with a larger number, 100 seems to be a "happy medium." Of the several lists extensively used, Jung's has been carefully prepared to locate common complexes and the Kent-Rosanoff list has been made up "to avoid such words as are especially liable to call up personal experiences."

In making up a list, only those words should be selected which are commonly used and are presumably understood by the sub-

ject with the least extensive education. Thorndike's Word List* will be of service here. Words of double meaning should not be employed. If the list is to be used for uncovering complexes, some of the words included must be irrelevant, innocent, or matter-of-fact words referring to things to which persons are normally well adjusted; while others may be relevant or critical words which are related to common complexes. These should be arranged in a chance order with the exception that the first ten words or so, if innocent, will assist in mechanizing the mode of reaction. In this connection it may be mentioned that Hull and Lugoff (28) found the words in Jung's list were arranged in a regular alternation of strong and weak. Accordingly if the Jung list is used, it is well to preface it by ten or twenty words for orientation purposes.

When the association experiment is given individually, there are two alternatives, one to present words visually for the subject to read, the other to present them orally. Since reading is a skill in which persons vary widely, it is found more satisfactory on the whole to give the words orally.

Some genius ought to set himself to devise a method of giving the association experiment in groups. As things are now, if the purpose of the test is merely to obtain the associations, it can be given in groups by having the stimulus words dictated and letting subjects write their reaction words. But since reaction time is one of the most important evidences that the experiment affords, much of the value of the method is lost when subjects merely write their responses.

Elaborate devices have been used in order to record the time accurately. Münsterberg (54) describes a set-up in which a little instrument is held between the lips of both the experimenter and the subject. With such an instrument the least movement of speaking makes or breaks an electric current passing through an electric clock-work whose index moves around a dial every second. When the experimenter moves his lips to give a word, he starts the pointer revolving. When the subject opens his lips to respond, the current is broken and the pointer stops. The association time can thus be measured to the thousandth part of a second.

*Thorndike, E. L., *The Teacher's Word Book*, 2d ed., Teachers College Bureau of Publications (1931).

Dunlap (11) has investigated the relative merits of the chronoscope and stop-watch in the measurement of reaction time. Although the stop-watch is relatively inaccurate and gives longer readings than the chronoscope (due to the lag in shutting off the watch in response to the subject's voice), it has some advantages over complicated apparatus which distracts the attention of the subject. Those who have considered fairly the merits of the two methods prefer the stop-watch.

Jung (34, p. 228) maintains that it is accurate enough to use a stop-watch registering fifths of a second, to be manipulated by the experimenter. Even though considerable error may result from the use of such a crude instrument, Jung believes that the error is still well within the limits needed by the method in its present stage of refinement. About five associations a minute can be taken and recorded using this method.

In giving the test, the instructions are to answer "as quickly as possible with the first word that comes to your mind" or "with the first thing this word makes you think of." The test should be given with the examiner and subject alone in a room free from distractions. It is well to have the subject seated comfortably where he cannot see the examiner, so that he will not be disturbed by the timing and the record.

The experimenter should record everything possible—the response, the reaction time, the general behavior, and all movements and expressions.

Wells,* in describing a blank for recording the report, suggests that the following symbols be used to indicate peculiarities in the response:

- c* Desires to change response word
- f* Fidgets
- i* Interjections not intended as response words
- l* Laughs
- m* Failure to understand stimulus word
- v* Repeats Stimulus word
- s* Speaks phrases or sentences which may contain response word.
- w* Gives more than one response word
- *-300 No response produced within one minute
- ? Asks questions about stimulus or response

*Wells, F. L., *Mental Tests in Clinical Practice* (World Book Company, 1927), p. 211.

366 Diagnosing Personality and Conduct

It has been found useful to go through the list a second time for "reproduction." Here one determines whether the subject remembers or is willing to repeat the response he made first. The instructions are: "I am going through the list of words once more, and I want you to try to give me the same words you gave me before. If you remember or think you remember, say it. If you don't, say 'No.' You can take as long a time as you want to answer."

Some writers, especially those of the psychoanalytic school, also require the subject to give the train of associations which each word arouses. They not only wish to have the subject respond with a word, but they wish to get at what is back of the association.

Before the method is discussed critically, mention should be made of the utter worthlessness from the experimental point of view of much of the work done on the association method. Workers in the psychoanalytic field, usually practitioners with the clinical point of view, are the victims of their enthusiasm. Their work has made them give close attention to individual cases, in which symptoms achieve a magnified importance. Their "experience," with all of its biases, overemphases, and neglects is placed before the results of strict scientific inquiry. Mere contiguous associations are elevated to the level of universal laws. In consequence of this, many of the facts reported concerning the association method in the following pages may later be overthrown when given more exhaustive experimental tests.

The greatest deficiency in this experimental work is the small number of cases deemed to be sufficient in drawing important conclusions, a deficiency to some extent pardonable, since each test must be given individually and the method is time-consuming. Jung actually draws important conclusions from his experiences with a single case. Much of the work has been done with fewer than ten cases. This might not be serious if the results were entirely consistent among the few cases tried, but they are not. Many and many a time in psychological experimentation a difference found with a few cases has dwindled down to practically nothing as their number has been increased.

A second inadequacy of much of the work done arises from a need for special insight and skill on the part of the experimenter, which is often unmet. A diagnostic method may properly require

skill on the part of the observer, but ought not to demand a sagacity not adapted to thorough description in a manual. When a method requires judgment, successful results from its use tend to become matters of choice—it becomes an art rather than a science.

A third error, not so much in the method itself as in the conclusions drawn from it, is connected with the matter of types. One of the easiest pitfalls into which the psychologist falls is that of dividing people into types. This error results from a failure to comprehend the principles of individual differences and the continuity of these differences. Those who are accustomed to working with large numbers of cases and with finely divided scales of measurement learn that whenever human traits are graded along a scale representing different amounts or degrees of the trait, the distribution practically always assumes the shape of the normal probability curve. But the clinical worker who sees only a few cases is unable to bridge gaps in the scale and falls into the easy explanation of types.

Wells (78), for example, in a certain experiment, divides his subjects into two groups, A and B; the former consisting, according to ratings, of persons who are intelligent, prompt, persistent, active, conscientious, truthful, deep of mood, sensible, and taciturn, while those in group B are intelligent, prompt, active, self-appreciative, loquacious, superficial in mood, and low in truthfulness, conscientiousness, and sensibility. He fails to note that there is no clear dividing line between these two groups, and also that there is a far from perfect correlation between the different traits which define a given group.

Perhaps the clearest case of the error of assuming the existence of types was made by Marston. In early work it had been found that deception was a state characterized by long reaction time. Marston (46) in later experimental work with only ten subjects found that in giving deceptive responses four had longer reaction times than normal, three had shorter reaction times than normal, and three showed no marked divergence from the normal. He concludes that there is a "positive" and "negative" type, quite overlooking the fact that individual differences may cause considerable divergence above and below the mean of the group, a mean probably slightly higher for subjects showing deception than for subjects working without deception.

Classification of Response

A large amount of work has been done in the classification of the responses in free association. Since answers are neither right nor wrong, and vary in many ways, some sort of grouping was found necessary before the replies could be interpreted; and though most of those classifications have proved sterile and will not be reviewed, there are some which show significant differences in personality, and others which help in interpreting differences in reaction time to be mentioned later.

Galton's (16) original classification is of interest. He divides all associations into three groups: (a) those that are predominantly *imagery*, (b) those that are *histrionic* and in which tendencies to react are uppermost, and (c) the *abstract*. Galton believed that the histrionic associations were most frequent with him. But individuals differ in this respect. Rusk (64) was amazed by the "astounding definiteness and vividness of children's imagery."

Another famous classification is that furnished by Jung (34, p. 38) in working over the previous classification of Kräpelin-Aschaffenburg. The primary classification is:

- A. Intrinsic (inner) associations
- B. Extrinsic (outer) associations
- C. Clang (sound) associations
- D. Miscellaneous associations.

In the group of intrinsic associations belong those which come by way of the meanings of words. One suggests another because of some common element in the meaning. For instance, *orange* suggests *apple*, both being fruit, or *water* suggests *wet*, a quality of water. In the group of extrinsic associations are those of contiguity, of mere contacts in space or time. *Pen—ink* is an example; or *bitter—sweet*. In the third group come those cruder associations of sound such as rime, or some even more elementary sound resemblance such as *make—shake* or *window—winter*. Into the fourth group are thrown the miscellaneous associations, chief among which are the *mediate* or *indirect* associations in which a middle term is necessary to complete the link.

This classification, expanded, is the basis of the following scheme taken from Jones's *Papers on Psychoanalysis* (29, p. 431).

A. *Intrinsic Association—Continuity.* An essential resemblance present between the meanings of the stimulus and reactions words.

- | | |
|---|---|
| I. Coördination. Essential similarity between the two. | Apple—pear |
| II. Predication. The reaction-word expresses some predicate, judgment, function, or attribute of the stimulus-word. | Snake—poisonous
Book—something to read |
| III. Causal dependence. The idea of causation implied in the response. | Pain—tears |

B. *Extrinsic Association—"Contiguity."* The resemblance present is a superficial or 'chance' one.

- | | |
|---|------------------------|
| I. Coexistence. Simultaneous. The two ideas connected through frequent simultaneous use. | Pen—ink |
| II. Identity. Synonyms or nearly so. | Effect—result |
| III. Motor-speech forms. The two words connected through frequent use in daily expression, proverbs, qualities. | Pen—Sword
Cat—mouse |

C. *Sound Association.* The resemblance between the two words being primarily an auditory one.

- | | |
|---------------------|------------|
| I. Word completions | One—wonder |
| II. Clang | Line—lying |
| III. Rime | Cast—past |

D. *Miscellaneous.*

- | | |
|--|---|
| I. Mediate. An indirect association, intelligible only on the assumption of an intermediary bond that does not appear in the reaction. The association of the bond may be any one of the forms mentioned above, and its relation to the stimulus word (centripetal) and to the reaction word (centrifugal) can be separately classified. | Run—rifle (centripetal sound disjunction, <i>gun</i> being the intermediary word) |
| II. Senseless. No discernible connection between the two words; in this case the reaction-word us- | |

370 Diagnosing Personality and Conduct

ually refers to some object in the immediate environment.

III. Failure. No reaction at all.

IV. Repetition of the stimulus word.

Still another classification for which Jung (34, p. 168) is responsible is as follows:

A. *Objective.*

I. Reaction principally conditioned
via the objective meaning of the stimulus word.

II. Reaction principally conditioned
via the linguistic features of the stimulus word.

B. *Subjective. Egocentric.*

I. Predicate.

a. Personal judgments. Emotional.

b. Definition. Intellectual.

II. Constellation.

a. Simple constellation.

b. Complex constellation.

Noble—man

Car—a vehicle for transportation.

A word of explanation for the two latter groups. By a *simple constellation association* is meant a reaction influenced by special individual complexes strongly invested with emotion. In the complex type the influences are "unconscious," that is "the person is not aware of their content, which, being too unpleasant to remember, has been buried."

Still another classification scheme is that of Kent-Rosanoff (37).

I. Common reactions

1. Specific reactions

2. Non-specific reactions

II. Doubtful reactions

III. Individual reactions

1. Normal reactions

2. Pathological reactions

A. Derivatives of stimulus words

B. Partial dissociation

(a) Non-specific reactions

(b) Sound reactions

i. Words

ii. Neologisms

- (c) Word complements
- (d) Particles of speech
- C. Complete dissociation.
 - (a) Perseveration
 - (b) Neologisms without sound relation
- 3. Unclassified

By a *common reaction* these writers mean a response that is given by large numbers of people; by an *individual reaction* is meant a response given infrequently. In an exhaustive study of the responses of 1,000 individuals, Kent and Rosanoff have drawn up tables of the frequency with which various responses occur, and these tables may be referred to for deciding whether a reaction is common or individual.

These classifications are of value only in helping one to interpret the significance of various responses, but to classify accurately a hundred responses is difficult if not impossible. Certain of the classification categories overlap, others are vague, and to assign a response to others would require insight or skill beyond the reach of any experimenter, since the mere association simply is not a guide to its own classification. In short, such a classification is at best extremely subjective.

Factors Conditioning the Response

Jung (34, p. 167) has shown that the type of reaction is a function of the quality of attention. When attention is heightened, the association is *meaningful*; when attention is relaxed, inhibitions are broken down, and verbal and clang reactions are given. Jung holds to the somewhat fanciful explanation that these clang reactions are infantile and require to be inhibited in normal purposeful activity. In corroboration, Aschaffenburg (38, p. 566) has found that under fatigue, associations become more superficial and the relation between the stimulus word and the reaction becomes looser. Wells (81) has shown that reactions become more superficial after practice, in effect also attributable to weariness or fatigue. Wells also finds that practice tends to differentiate and particularize the response. A similar tendency toward superficiality of association has been noted as the result of the action of drugs, such as alcohol, tea, coffee, etc. In short, as the attention is relaxed, the quality of the association lowers.

372 Diagnosing Personality and Conduct

Jung (34, p. 168) again tells us that the educated (intelligent?) also exhibit this distraction or dissociation phenomenon, and that their associations are more shallow than those of the uneducated. He explains this (in part) on the ground that the uneducated person lets himself go much less and allows fewer disguised subjective wishes and valuations to break through (p. 130) a psychoanalytic explanation which seems the longest way round. A much simpler and more direct explanation given by Jung is that the uneducated man responds to words as fragments of sentences and according to the experience aroused. Words are vicegerents of the situations they represent. The uneducated person responds to a word subjectively, personally, as though he were being asked a question. The educated person, on the contrary, is used to considering words by themselves, and is therefore ready to give a merely verbal association.

Allport (2) finds fewer egocentric reactions when one is working in a group than when working alone. Apparently the group stimulates "group" reactions, while inhibiting reactions that are strictly personal or egocentric.

Woodrow and Lowell (88) show that children show fewer individual reactions than adults, using tables of children's frequency of response. This contradicts findings by Rosanoff and Rosanoff (63), who interpreted the children's reactions on the basis of their adult tables. A simple interpretation of the findings of Woodrow and Lowell is that children have a smaller stock of words on which to draw and fewer peculiar or individual experiences. They have not yet, as a rule, met the variety which results from the vicissitudes of living. Some of the earlier writers who believed that children gave a larger number of individual responses tried to explain this by asserting that children were more proof against "partial dissociation." This is only another illustration of the ease with which imagination tends to run ahead of the facts in this field.

There seems to be a difference in testimony, also, as to the *reliability* of these classifications. Rosanoff, Martin, and Rosanoff (62) report that there is great variability when the test is repeated after an interval of three minutes. On the other hand, Guthrie reports a reliability coefficient of .80 (one half against the other) when the free association test is scored by the number of "common" (Kent-Rosanoff) responses.

Some of the major factors conditioning the response have been mentioned, including variations in the attention, education, and age of the subject. Other major factors undoubtedly exist, but in the main one must look to the individual experience for an explanation of the precise word given.

Reaction Time and Factors Related to It

Many investigators believe that the reaction time is more significant than the actual reaction itself. By *reaction time* is meant the interval between the instant the stimulus word is presented and the instant the response word is given by the subject.

1. The first factor to be considered is that each individual has a reaction time that is peculiar to himself. Differences in reaction time were investigated by Cattell (6) long ago, and the facts are well known to-day. On the other hand, it would be cumbersome to have to obtain an individual's reaction time tendency before using him in the association experiment. In practice it is common to assume a median reaction time of two seconds (Jung, 34, p. 231, gives 1.8 seconds) measuring deviations from the normal from this figure. Assuming that individuals vary in their normal reaction time around this median value, let us inquire what other factors influence reaction time.

2. It has been noted that "free" association has a longer reaction time than "controlled" association. Ask a person to give the opposite of *hot*, and the response will come more quickly than if you give him the word *hot* and ask him to tell you the first word he thinks of. When the additional control "opposite" is given to the subject, the range of alternatives is vastly decreased, but when one is given freedom to utter any word whatever, a time-consuming choice must be made from many possibilities. If for any reason the choice is difficult, or if the first impulse to speak a word is blocked, the reaction time may be considerably lengthened. It is this that makes reaction time so serviceable in the diagnosis of complexes.

3. Crane (9) finds that when the word is presented visually, the reaction time is longer than when it is presented orally.

4. Woodworth-Wells (87) found that the reaction time to the first word in a list is considerably slower than to subsequent words in the list or to words given in isolation.

5. Jung (34, p. 234 ff.) finds that the type of stimulus word has its effect on the reaction time. The shortest reaction time followed the presentation of concrete words (1.67 seconds); the longest reaction time followed abstract words (1.95); while verbs occupied an intermediate position (1.9). Somewhat the same thing was found by Crane (9), who reports that adjectives yield the shortest times, nouns next longest, and verbs the longest of all. Probably there are three factors at work here. One is that we are more familiar with concrete than abstract words, and therefore the associations are more plentiful and more ready. A second factor is that adjectives and nouns usually require nouns and verbs respectively for completion, whereas verbs do not regularly require words for completion of the thought. Hence it is relatively more difficult to obtain responses to verbs. Still a third may be that the response to nouns and adjectives tends to be objective and impersonal on the whole, whereas the response to verbs tends to be subjective.

6. Jung (34, p. 236) points out that there are also differences in reaction time according to the kind of response. Names of abstract concepts require the longest time to produce (1.98 seconds), concrete words are next (1.81), while adjectives and verbs come most easily (1.65 and 1.66). He also reports that the educated take the longest time in giving concrete words, but this conclusion must be accepted with reserve inasmuch as it is based on only a few cases.

7. With regard to the reaction times for the different logical categories used in classification schemes, Wells (84, p. 22), after several pages of discussion, concludes that "In the Kent-Rosanoff experiments there was evident no general tendency of the more frequent classes of associations to be shorter or longer than the less frequent ones."

This, perhaps, is because the Kent-Rosanoff classification scheme is not very significant. Jung (34, p. 238) reports that *inner* associations (according to the meanings) are longer than *outer* associations (the more superficial and verbal), while the clang associations have the longest reaction time of all.

8. Allport (2) found that the speed of reaction was greater (reaction time less) when the subjects worked in groups than when they worked alone.

9. Wells (81), who has made an extensive study of the effect of

practice on free association, reports that reaction time decreases after practice.

10. Although some of the early investigators (Ziehen especially) reported that the speed of association increases with age, this is denied by Rusk (64), who finds no definite trend of this kind in his work.

11. Alcohol and caffeine cause a slight decrease in reaction time, according to Langfeld (41).

12. Jung (34, p. 231) points out that men have a shorter reaction time (1.6 seconds) than women (2.9 seconds). This phenomenon has been noticed by several observers and may be regarded as well founded. But the explanations are many and diverse. A common explanation is that women are more emotional than men. Wells reports in this connection that women are more variable than men in their reaction time.

13. Jung (34, p. 232) again reports that educated (intelligent?) subjects show a shorter reaction time than uneducated subjects.

14. Ekdahl (13), in a recent study, presents evidence to show that the reaction time is influenced by the presence of the investigator. He found that when the association experiment was conducted by mechanical aids without the experimenter, the reaction time was lower than when the experimenter was present. Part of this may have been a confusion or inhibition due to the presence of another person. It is well known that a skilful experimenter by intonation can give words a more personal appeal, arouse complexes, and lengthen reaction time.

15. Words that carry unpleasant sense qualities lengthen reaction time, according to Tolman and I. Johnson (75) and also W. Smith (67, p. 249). On the other hand, Rusk finds that there are no significant differences in reaction time between pleasant and unpleasant words for children. The fact that the affective tone of words influences reaction time seems pretty well attested for adults; Tolman and I. Johnson find that women are more prone than men to lengthen responses to unpleasant stimulus words and that, in the case of women, there is a decrease in reaction time for words pleasantly tinged, a fact which perhaps helps to explain the greater variability of women.

There is a certain consistency in all of these findings that the reader must trace out for himself. When it is considered that reactions are more often superficial when given by men than by

women, when given under fatigue, under drugs, when given by educated persons, and after practice; and that the reaction time is shorter for *outer* or superficial associations, after practice, under the influence of drugs, for men, and for the educated, it becomes clear that reaction time is not a matter of chance but is under the influence of one or more very definite forces. Not all of these factors have been demonstrated with finality, since in some cases the experimental evidence is contradictory or confusing. However, a general consistency apparent throughout the work makes it appear probable that future investigations will confirm most of the relationships noted above.

It should be emphasized repeatedly that none of these influences on reaction time is large, and that there is great overlapping. For example, although we may say that men have shorter reaction times than women, this does not mean that all men have shorter reaction times than all women. Since there is much overlapping between the sexes, the difference in the averages is very small, and the biserial r between sex and reaction time is very small.

After all of these factors (each of which is small and distinct, though on the whole relatively unimportant) has been taken into account, large differences in reaction time will be found left over. Where the influence of the foregoing factors would be manifest in times of tenths of a second, reaction times frequently occur which are several seconds in length. It is believed that such lengthened reaction times are caused by a "complex" or emotional blocking, as when the word or the situation which the word represents evokes an emotional response, such a host of ideas surge in that the subject finds it difficult to make any response. In most of the cases the subject tends to hide or conceal his first impulse to respond as being something too personal, too revealing, or too disagreeable.

The evidence that lengthened reaction times are due primarily to the influence of emotion was derived originally from clinical investigations in which Jung and those who preceded him discovered that there was a relation between the words showing lengthened reaction time and the complexes of their patients. More recently there has been striking confirmation of this through experiment. Peterson and Jung (58), and W. Smith (68, p. 77) report that there is a marked correlation between galvanometric deflections and reaction time in free association. As may be noted

from the chapter on "Physiological Measures," the galvanometric experiment is a well-established method for detecting the presence of emotion. Smith finds the correlation between the psychogalvanic reflex and word association reaction time to be .47. He believes the power of the word association method is increased if it is used in conjunction with the psychogalvanic reflex. As a measure of emotion the galvanic reflex is superior to reaction time. In this connection reference should be made to the work of Landis, Gullette, and Jacobsen (39), who found that reaction time was one of the best criteria of emotionality available.

Another bit of evidence is the inspection of the words themselves that show long reaction time. In most cases the emotional character of the words is clearly evident. Jung finds the following words to give unusually long reaction times: *needle, hop, strange, false, heart, pyramid, strike, threaten, remember, ripe, to woo, fern, hair, nauseate, dream, paper, book, harm, softly, caress, family, consciousness, freedom, faith, violence, wonder*. Wells (84, p. 13), using the Kent-Rosanoff series, finds the words yielding the longest reaction times to be: *square, command, anger, religion, moon, slow, health, justice, stomach, wish*. W. Smith's (68, p. 75) list of words yielding long reaction times is: *name, friend, despise, make, sell, proud, home, nasty, marry, habit, pity, happy, angry, bring, dance, worry, kiss, brother, family*.

W. Smith (68) also tried out the reproduction of response words and the memory for them after a considerable lapse of time. He finds that words with high affectivity are best remembered, but that words soonest forgotten have an appreciably higher affective value than words which are moderately well remembered. From this he infers that there are two distinct varieties of affective tone which he calls *positive* and *negative*, which correspond somewhat to *pleasant* and *unpleasant*. He finds that prolongation of reaction time is, in general, a sign of *negative* or *unpleasant* affective tone.

In this connection Smith (68, ch. V) believes that the word association method may be used for determining the integration of personality. He finds, for instance, that the reaction time for the list of words for a given individual tends to correlate .65 with a repetition of the list for that individual. The average correlation of two individuals is about +.15. If an individual is consistent in his responses, his personality is highly unified, whereas if he is inconsistent there is evidence of instability or dissociation.

Wells (81) points out that the value of the association experiment for measuring the affectivity of the subject decreases with practice.

In summary the evidence is conclusive that the reaction time of word association is a valuable measure of the affectivity of a subject in general and to certain situations. Although the word association reaction time is first a function of the reaction time of the individual and second is influenced to a slight degree by a variety of miscellaneous factors such as sex, nature of the stimulus, nature of the response, age, education, and setting of the experiment, large variations in reaction time are in great measure due to the presence of emotion.

Complex Signs

Now that the value of lengthened reaction time has been discussed as an indication of the presence of emotion or a "complex," let us turn to the general question of the value of the association experiment for uncovering situations that arouse emotional responses. This has been given considerable study by the earlier investigators, who have enumerated a long list of "signs." Whereas the earlier investigators discarded any word from the list to which the response was irregular, the later ones found that these irregularities and aberrations of response were extremely significant. Some of the more important complex signs are as follows:

1. *Long reaction time.* This is the only quantitative measure available. Any reaction requiring over 2.6 of a second is usually considered significant.

2. *Inability to make any response whatever.* Occasionally no response will be elicited even though time up to a minute is allowed. Such failure to respond may be due to a number of factors, among which may be mentioned inhibition of any response; articulatory block; attention diverted by copious or diverting imagery; absorption in trains of imagery or reverie; competition of reaction words; or no meaning found in the stimulus.

3. *Extremely short reaction time.*

4. *Repetition of the stimulus word itself.*

5. *Apparent misunderstanding of the stimulus word.* The psychoanalytic explanation is that in such cases there is a strong

desire not to understand. But this explanation need not be assumed. In some cases it may well be true that there was a definite misunderstanding due to faulty learning, or indistinct or strange pronunciation on the part of the examiner. Perhaps in all such cases the prepotency of any part of the word is also influenced by competing ideas or images perseverating from previous associations.

6. *Defective reproduction of original reaction at second presentation of the stimulus word.* In the reproduction experiment, if the second response differs from the first, suspicion of a source of irritation arises. Smith (68, p. 78) believes this to be the best complex indicator.*

7. *Response with the same reaction word to two or more different stimulus words.* This is sometimes called perseveration. Perseveration may be due to a certain complex or constellation which dominates consciousness, or to a poverty of ideas or to other more significant causes. In some cases where the subject suspects the nature of the experiment he may avail himself of perseveration to assist in concealment.

8. *Strange or apparently senseless reaction.*

9. *Perseveration of ideas.* In this case, though the exact word may not be repeated as in number 7, the same idea perseverates in two more responses.

A variety of other complex signs have been noted, especially peculiarities in the response and uneasiness in the behavior of the subject. Even when the response itself seems to have peculiar significance, unless the experimenter is confident of his judgment he should be alert to prevent his imagination from imparting significance where it does not really exist. Suspicion of emotional irritation may be aroused by such mannerisms as whispering the response, bodily movements, nervous movements of the hands, reddening of the face, coughing, clearing the throat. In all probability many valid physiological indices could be determined by laboratory instruments such as changes in the pulse, blood-pressure, breathing,** knee-jerk, strength of grip, etc. Giving more

* In this connection Peterson and Jung⁽⁶⁸⁾ report that in correlating responses in the reproduction experiment with the psychogalvanic reflex they found altered reproduction associated with increased deflection of the galvanometric needle.

** Nunberg, who has studied variations in breathing under emotion, believes that "unconscious" complexes show strong inhibition of respiration, whereac

380 Diagnosing Personality and Conduct

than one word in response is a very suspicious sign.* Kohs (38) lists such types of response as quotations, titles, sentences, addition of the article, example (*the person*), or responses in a foreign language, as being significant. A familiar subterfuge which may betray the presence of a complex is the naming of some article in the examiner's room.

Thanks to the work of Hull and Lugoff (28), we have an experimental check on several of these "signs" as indicators of a complex. They obtained responses to a list of 100 words from fifty men and fifty women.** The occurrence of the first nine of the complex signs listed above was noticed. Correlations were computed between each of the "signs" and also between each sign and a composite of all signs grouped together. These correlations are shown in the accompanying table.

TABLE 77
CORRELATIONS BETWEEN COMPLEX SIGNS IN FREE ASSOCIATION
(from Hull and Lugoff, 28 p. 121)

	<i>Av. no. of appear- ances per subject</i>	<i>I</i>	<i>II</i>	<i>III</i>	<i>IV</i>	<i>V</i>	<i>Com- posite minus V</i>
I. Repetition of the stimulus word....	4.53		+.41	+.69	+.14	+.24	+.53
II. Misunderstanding.	1.02	+.41		+.52	+.06	+.31	+.33
III. Long reaction time	20.38	+.69	+.52		+.26	-.09	+.24
IV. Defective reproduction	19.14	+.14	+.06	+.26		+.02	+.17
V. Repeated use of the same word....	20.15	+.24	-.31	-.09	+.02		-.06

From this Hull and Lugoff conclude that (a) repetition of the stimulus word is decidedly the most reliable diagnostic sign of

with "conscious" complexes there is excitation in addition to the inhibition. Nunberg, H., "Über Körperliche Begleiterscheinungen Assoziativer Vorgänge," *J. f. Psychol. u. Neurol.*, 16: 102-128 (1910).

* In this connection Jung (34, p. 202) remarks that the feeble-minded seldom react with one word and those with low mentality have a tendency to give phrases or sentences rather than a single word. Often the tendency is to give a definition of the stimulus word. Accordingly one must not interpret the tendency to give more than one word as a symptom of the presence of a complex unless other signs give corroboration.

** No significant sex differences were found. Certain words differed markedly in the number of signs yielded by men and women, but on the whole the men and women showed about the same number of signs on the various words.

the five indicators examined; (b) the first four given in the table are in all probability real complex signs; (c) repeated use of the same reaction word is a complex sign of very little value.

As to extremely short reaction times, Hull and Lugoff find that they are neutral in indicating the presence of a complex.

Finally, Hull and Lugoff studied experimentally a point previously emphasized in the literature that one should not depend on the occurrence of *one* of the signs for deciding the presence of a complex. The earlier users of the method concluded, "One may enter a judgment, only when in possession of an *aggregate* of indicators." Different individuals probably betray their complexes in different ways. But Hull and Lugoff believe that the increase in diagnostic potency resulting from the addition of successive indicators follows a law of diminishing returns. "While two indicators are distinctly more significant than one, the second indicator adds by no means as much diagnostic potency as the first."

One of the phenomena stressed by Jung is the tendency for the emotional tone of a stimulated complex to persist in the next few responses, i.e., to persevere, so that the tendency to make a long reaction time to a certain word might last over the two or three following words which were in themselves innocuous. It has been observed that a word containing considerable affect may evoke a response having a normal reaction time, but that the affect persists and may cause long reaction times in the next few words. For instance *lose* may immediately arouse the memory of clothes lost in a fire, and the response *clothes* may be given with no increase in reaction time. But the train of thought aroused may persist and interfere with normal response to the next word or two. Hubbard (27) studied this experimentally and found that this observation had no justification in fact, that the perseveration tendency, if such there be, is too small to be of diagnostic significance.

In the same experiment Hubbard discovered the important fact that the significance of a complex sign varied according to its position in the series of words. For instance, long reaction times are more significant at the beginning or end of the series because reaction times normally tend to be longer in the middle of the series. Again the number of individual responses (in the Kent-Rosanoff sense) increases as one advances further in the series.

The Free Association Experiment as a Test of Lying and Guilt

It was Münsterberg (54) who first studied the use of the free association experiment for the detection of lying and guilt. In a series of popular articles he gives a vivid picture of the possibilities of the method in catching and convicting criminals and predicts the use of it as an adjunct in the operation of the law. Though certain difficulties have prevented its practical application, the principle underlying it is secure, and it may confidently be said to await only the skilful technician to put it to legitimate and practical use.

A description of the method, as worked up into a clever and startling laboratory exercise, provides perhaps the clearest exposition of its nature. As Münsterberg pointed out, a criminal who is hiding something has to suppress reactions of an emotional nature. This inhibition will lead to an increased reaction time to critical words presented for response. On the other hand, a hardened criminal who has nothing to conceal shows no emotion and hence makes no betraying reactions in the association experiment.

The following description of a class experiment in the detection of suppressed ideas is quoted from the *Elementary Laboratory Course in Psychology* by Langfeld and Allport (42, pp. 112, 113).

"Two subjects are selected, and are given a sealed envelope containing instructions for the crime. The two subjects leave the room and decide between themselves which is to play the criminal. The 'innocent' subject waits in the hall or in an adjoining room until he is called.

"The 'criminal' now opens the envelope and follows the instructions given. It is important that the innocent person should know nothing about the crime, and that no one but the two subjects should know which one is guilty. The crime, which should be arranged beforehand by the experimenter, should preferably be one which arouses a certain degree of 'criminal consciousness.' The following is given as a model for the instruction in the envelope.

"Go into room —, and there you will find in the desk drawer a sealed letter addressed to one of the members of the class. Open it, tearing the envelope as little as possible, and read the contents carefully. Then replace the letter in the envelope and, with the aid of the mucilage you will find on the desk, reseal the flap of the envelope neatly so as to avoid detection. Then put the letter back in its original place in the drawer.

"When you are called into the class-room for the experiment, use all your ingenuity to conceal your guilty knowledge of the crime."

"In a crime of this sort the letter should reveal some very personal, though fictitious affair of the student to whom it was addressed, such as financial difficulties or a paternal reprimand for debts and extravagance, etc.

"After the instructions have been carried out, the 'criminal' joins the 'innocent' subject and awaits summons from the 'court.'

"In the meantime the experimenter informs the class of the details of the crime. A list of fifty common words has been prepared beforehand by the experimenter. Of these words twenty-five are closely related to the details of the crime. Such words are called 'crucial words.' The crucial words should be distributed throughout the list, some occurring singly, others in groups of from three to five. A partial list is given below as an illustration, supposing the crime to have been the one described (though actually a totally different crime must be prepared):

chair	disk	church
hat	black	fork
horse	supper	coffee
man	broom	<i>mucilage</i>
dinner	pencil	<i>drawer</i>
keys	<i>letter</i>	<i>money</i>
watch	<i>open</i>	<i>expense</i>
window	<i>debts</i>	etc.
<i>envelope</i>	tree	
telephone	<i>school</i>	

"One of the subjects is now called in and seated facing the class. The experimenter speaks the first word of the list distinctly, and the subject replies as quickly as possible with the first word that comes into his mind."

A number of experiments in the detection of guilt have been reported in the literature. In general the results show that with naïve subjects unacquainted with the method, the success of detection is remarkable. For instance, Leach and Washburn (44) tried out the method on twenty-six subjects with two judges, who studied the results independently and made only one error in the fifty-two judgments. Langfeld (40) in an experiment on two subjects, one "guilty" and one "innocent," reports that the innocent subject showed reaction times to the crucial words averaging .37 of a second longer than those to the non-crucial words, while the guilty subject showed a difference in the same direction amount-

ing to .83 of a second. The delay for the guilty subject was over twice that for the innocent subject.

The question always arises whether in this experiment one should use reaction time as the sole criterion of guilt, or include also a qualitative analysis of the responses to the critical words. The general consensus of opinion is that the qualitative analysis is so unreliable as to be practically worthless. Often an innocent subject will give a word response that might be interpreted as showing evidence of crime merely because that response is a common one to that stimulus word. On the other hand, guilty subjects will frequently misinterpret a critical word and give an innocuous response. Coupled with this is the tendency on the part of the examiner to let his imagination construct hypothetical relationships in no wise responsible for the reaction. The only safe course is to rely strictly on reaction time.

Reaction time would yield overwhelming evidence were it not for the fact that in any innocent individual a given list of words will touch off "private" complexes not in the least connected with the particular event or "crime" being investigated. This cannot be avoided. But by the laws of chance there should be as many of these private complexes evoked by the crucial words as by the non-crucial words, and for both guilty and innocent subjects. Accordingly the difference in direction in response between the crucial and non-crucial words should not be affected by these extraneous influences, and the only change should be in the direction of making the amount of difference smaller.

The stumblingblock to the successful use of the method in detecting crime in the actual situation, over and above the objection previously mentioned that some criminals are concealing nothing emotionally, is that with a knowledge of the method it is possible to conceal deception. In experimental work with naïve subjects the method has been uniformly successful, but with sophisticated subjects the results are by no means so unanimously successful. Steinberg reports that out of twenty-three tests with sophisticated subjects fourteen correct and nine incorrect judgments were made using the criterion of larger *average reaction-time* for crucial words, but that only nine correct and fourteen incorrect judgments were made using as a criterion the differences in *average deviation* of reaction time of crucial and non-crucial words.

There are various devices that a sophisticated subject can employ to conceal his deceit. He may give as his response some word suggested by a previous word in the list or some previous response of his own, or a list of cities or a list of names or objects in the examining room. Or the subject may have responses prepared in advance to words that he guesses will be used as crucial stimulus words. Since crucial words are often names or objects at the scene of the crime, preparation of response words to these objects might easily be attended to ahead of time. There is some doubt, however, whether this subterfuge would be uniformly successful. However, such devices, although they might control the reaction time to crucial words, ought to become immediately evident to the examiner, who could see that the subject was dodging the requirements of the experiment.

Because of the possibility that a person conversant with the association method may be able to conceal his guilt, the method is not at the present time a sure-fire technique. In the hands of a skilled examiner it probably could always be used on a naïve subject with successful results. It is probable, also, that a clever detective, willing to follow up cues, might make successful use of the association method even where the subject tried to conceal his acquaintance with the criminal situation. But the method cannot at present be called entirely dependable in the detection of lying and guilt.

The Association Method and Insanity

To Kent and Rosanoff (37) belongs the credit for investigating the possibilities of the free association experiment in the detection of insanity. Their approach is somewhat different from that discussed in the foregoing pages. Considering insanity as idiosyncrasy, these investigators worked on the assumption that mental abnormality could be discovered by noting the number of *unusual* associations given to a list of words. Accordingly their first task was to compile the frequency of association of 1,000 normal subjects for 100 words. These have been presented in a set of frequency tables, one for each stimulus word, showing number of times each response word is given by the 1,000 subjects. A *common reaction* is defined as one to be found in the tables, and an *individual* reaction as one which is not found in the tables. Any

386 Diagnosing Personality and Conduct

reaction word which, though not in the table in its identical form, is a grammatical variant of a word found there, may be classed as *doubtful*.

Using the same list on 247 insane subjects, these authors were able to draw up the following table.

TABLE 78

PERCENTAGE OF NORMAL AND INSANE SUBJECTS SHOWING COMMON, DOUBTFUL, AND INDIVIDUAL RESPONSES IN FREE ASSOCIATION

	Common reactions	Doubtful reactions	Individual reactions
1,000 normal subjects	91.7%	1.5%	6.8%
247 insane subjects	70.7%	2.5%	26.8%

This corresponds to a coefficient of contingency of .26 (uncorrected) or .48 (corrected), and indicates that the method has some diagnostic power. Mateer, who has used the Kent-Rosanoff list in clinical practice, suggests the following standards (47, p. 289):

"If a child gives more than 10 individual reactions on the Kent-Rosanoff association test and is more than eight years old, or if he gives more than 45 out of the most common reactions and is, in this latter instance, of normal level in intelligence, it may be taken as an indication of psychopathy.

"When the analyzed Kent-Rosanoff associations test is studied for quality, the test may be counted psychopathic in its indications if more than 10 reactions are found which are abnormal according to the Kent-Rosanoff definition of abnormality; or if they give that many indications of perseveration, automatism, sound association, repetition of the stimulus word, etc."

These workers went further and attempted to use the association method for diagnosing special types of insanity. First they drew up a scheme for the classification of responses (described on p. 370) and studied the differences in response made by different insane groups. Their table, which is too bulky to include here, reveals several significant differences in type of response. They find, for instance, that in dementia præcox there is a tendency to (a) give neologisms, particularly those of the senseless type; (b) to give unclassified reactions, largely of the incoherent type; and (c) toward stereotypy, manifested chiefly by abnormally frequent repetitions of the same reaction.* Other types of insanity do not

* Peterson and Jung (58) report no appreciable lengthening of reaction time in dementia præcox, but since this conclusion is based on the observation of only two cases, it may be dismissed as unproved.

show such clear-cut distinctions. The paranoiac cases are far from homogeneous in the types of responses which they give. Epileptics show tendencies to repeat one word or another many times, and an abnormally pronounced tendency to make use of non-specific reactions or particles of speech. Cases of paresis when mild give normal reactions; when severe, there is a tendency to perseveration. In cases of manic-depressive insanity there is again no uniformity in type of response.

One must confess that the work of Kent-Rosanoff, although representing an immense amount of industry, is decidedly faulty from the statistical point of view. Like so many investigators with training in medical science, they have a tendency to magnify individual observations and to fail to note inconsistencies. In the first place, as admitted by them, the method of classification of responses has a large element of subjectivity, a factor which cannot be easily dismissed and which remains as a cause of error in the results. Again, the inconsistencies and overlapping of response between different groups make diagnosis by this method alone very dubious. In the reports are records from normal subjects who make many perseverations, or unclassified reactions, or sound reactions, or non-specific reactions, and give every evidence of being abnormal. There is no sharp line of division between the sane and insane, but on the contrary a very gradual overlapping. The method, therefore, is presumptive or indicative only and cannot in its present stage safely be used alone in the diagnosis of insanity.

Free Association as a Measure of Ability

It is strange that the association method should have been considered as a measure of intelligence or of school achievement. Superficially there would seem to be no relation between the type of reaction given by the intelligent and non-intelligent, except that the intelligent person with a larger vocabulary ought to give a greater number of individual words. That this is the case has already been mentioned in this chapter. Jung points out the similarity between the type of response given by educated persons and by those suffering from various degrees of dissociation. There is probably a very different set of causes lying behind this similarity in type of response.

Various workers, however, have noticed qualitative differences in the free association response among the different grades of feeble-mindedness. Wreschner (89) found that with the idiot the quality of the stimulus word has a great effect on the response. The simpler the stimulus word, the higher the quality of the reaction word and the shorter the reaction time; any increase in difficulty or abstractness of the stimulus word is accompanied by a more superficial response and a longer reaction time. Wehrlin (ch. III, 34) finds that imbeciles seldom react with a single word, but receive the stimulus word as a question to be responded to with an answer. There is also a marked *tendency to definition* in the imbecile. He tries to explain the stimulus word instead of responding with the first word coming to his mind. Thus we find a shift from the characteristic response of a phrase or sentence that tries to fathom the meaning of the word among the feeble-minded up to the other extreme of mere verbal play on words, more or less superficial, among the highly intelligent.

Rosanoff and Rosanoff (63), who studied the responses of children to the Kent-Rosanoff list, report a definite correlation between association and mental capacity and also between association and school grade. However, their tables present differences in type of response between the bright and dull groups and also between pupils in higher and lower grades which are so slight as probably to be statistically insignificant. This difference cannot be computed, however, since the numbers in each group are not reported.

Eastman and Rosanoff (12) present an elaborate study of the use of the Kent-Rosanoff list on feeble-minded children in which they conclude "states of arrested mental development present certain fairly characteristic associational tendencies, characterized mainly by failures of reaction, non-specific reactions, and certain types of individual reactions." But here again the statistical technique employed fails to show the significance of the differences found or the amount of overlapping with normal children.

T. L. Kelley (36) has used the free association method with a technique of his own devising. In an early work he had studied the correlations of the association method with class standings in mathematics, science, and foreign languages. In the later experiment referred to he applied the Wells logical categories for classification, which have never yielded very significant results. Since

there were only twelve subjects, the probable errors of the coefficients of correlation were large, though some of the coefficients turned out to be large enough to be statistically significant. But the relationships are difficult to interpret. Altogether, the method is so cumbersome and difficult to use for the prediction of school achievement, and the results are so unpromising, that this work has never been followed up, and probably with reason.

Free Association as a Measure of Interest

Later in connection with the California Survey of Gifted Children, Kelley encouraged Mrs. J. B. Wyman (74) to use the association method in studying the *intellectual*, *social*, and *activity interests* of children. The treatment of the data in this work is unique and deserves a brief description. After some preliminary experimentation two lists of sixty words each were drawn up, the words being selected as "adapted to provoke responses due to intellectual interests, social interests, or activity interests." Children in the sixth and seventh grades gave their associations as in a group test by writing down the first word which occurred to them. These children were rated in various ways by their teachers for intellectual, social and activity interests. Finally groups of from fifty-eight to seventy-one children were selected as possessing or not possessing the different kinds of interest in question, though it was recognized that the two extreme groups in each case were merely the ends of a continuous distribution of the interests in question.

Then each separate response word was studied with respect to the frequency with which it was given by each of the contrasted interest groups (74, p. 463).

"A particular response word given by a large percentage of the 'with' group and by a small percentage of the 'without' group (or vice versa) has a high differentiating value and is diagnostic. For each response, therefore, the frequency was found separately for the 'with' group and the 'without' group, in terms of the percent who gave it. The difference between these two percents was also found. Next the S. D. of each percent was computed, and the S. D. of the difference between the two percents. Comparing the differences between the percents with the S. D. of the difference gave a measure of the amount of a particular kind of interest involved in a particular response word. Accordingly, the score

390 Diagnosing Personality and Conduct

assigned to a response word was the difference between the percents of 'with' and 'without' groups giving it, divided by the S. D. of this difference. The ratios thus obtained were transmitted into a 0-20 scale, 0 indicating no interest and 20 maximum interest. It was necessary, of course, to carry through this procedure separately for the three kinds of interests. In all, there were upward of 10,880 response words to evaluate, each three times. Upward of 13,000 additional responses were encountered in scoring the papers of other groups of children tested." *

Because many of the response words were similar in meaning or were merely grammatical variations, and also because certain of the response words were not given by one or the other group, making impossible the statistical treatment mentioned above, it was found necessary to fall back on certain arbitrary and subjective groupings in devising a scoring key. This key for a single stimulus word fills an entire page. Following is a brief section of the key for the stimulus word *gem*.

SCORES FOR RESPONSES TO STIMULUS WORD *Gem*

<i>Frequency</i>	<i>Responses</i>	<i>I.</i>	<i>S.</i>	<i>A.</i>
44		4	4	8
114	diamond	20	11	15
79	stone	12	10	15
57	jewel(s)	12	13	5

etc., for 72 other responses.

This means that when a given paper is being scored for *intellectual* interest, and the response to the word *gem* is given as *diamond*, a credit of 20 is given. The sum of the credits to the sixty words is a measure of *intellectual* interest. Each paper must be scored again for activity interest and social interest.

The method is sound statistically, but extremely laborious in the effort required to obtain the key in the first place and also to use the key to score a set of papers. Laslett (43) has used the same technique in constructing a free association test of *delinquency*, but it can be doubted that a method so very long and tedious will ever receive wide use.

Freyd (15) used the Kent-Rosanoff series in his study of the mechanically and socially inclined. He found that there were certain significant differences between the groups; for instance, the

* From Terman, *Genetic Studies of Genius*, Vol. I, 2d ed. (Stanford University Press, 1925), 463. Used by permission.

mechanically inclined gave more *individual* responses, while the socially inclined gave more *common* responses.

In conclusion it may be helpful to mention certain variations of the method, given by Kohs (38, pp. 568-572).

His first two methods consist of a discussion of the free association method and the reproduction (both of the response and the stimulus) method. Since they have already been discussed, Kohs' remarks on them will not be repeated here. Concerning variations on the reproduction method, he says:

"In these experiments, the examiner assumes that the subject has had a specific, vivid, emotionally-toned experience. He is then read a story ('Versuchsgeschichte') so constructed by the experimenter as to be very similar in content to the experience in question. The details in the narrative, however, are incomplete, vague, and only cleverly approximate to the actual events. Having been instructed to pay strict attention, the subject is requested to recall as much of the story as possible. But the details in the narrative and experience being so much alike, he more or less unconsciously confuses those of one with those of the other, and, as a result, inserts and supplements items which were wholly absent from the story. By this intimate knowledge of details, he easily betrays his familiarity with certain events. He may have previously denied most emphatically the least acquaintance with a single incident. Nevertheless, his responses definitely point to the contrary.

III. The 'Aussage' or Reproduction Method (Wertheimer)

"This procedure is not to be confused with Jung's. The individual here is asked to reproduce orally the story which has been read to him. Using the technical terms of the 'Aussage' Test, this portion of the method is the 'Bericht' or report, and may be further supplemented or completed by the 'Verhor' (the interrogatory or deposition). The latter may be any one or a combination of these three types. First, the questions may relate to details only, which are *common* to both story and 'Tatbestand' (experience, or constellation of facts); second, the questions may yield different responses owing to *differences* in story and 'Tatbestand'; third, the questions may relate to details present only in the 'Tatbestand.' Kramer states that this method may work perfectly in the laboratory, but may fail absolutely in actual application. More data are necessary.

IV. The 'Aussage'-Association Method (Wertheimer)

"Here the subject is given an association series made up of words of these three types: first, those relating to similarities in

both the 'Tatbestand' and the story; second, those relating neither to the 'Tatbestand' nor the story; third, those relating only to the 'Tatbestand.' The subject is requested to state his associations only to those words which appeared in the story; to all others he is to respond with the word 'nothing.' According to Lipmann the guilty easily betray themselves.

V. The Combination Method (Lipmann and Wertheimer)

"The following is Lipmann and Wertheimer's adaptation of the Ebbinghaus 'Kombinationsmethode.' After the subject has heard the 'Versuchsgeschichte,' a sheet is placed before him containing the text of the story, but from which many details have been omitted. Room is left in the 'Kombinationstext' for the proper information to be inserted. This information (not having been given at all in the story) either must be supplied from the 'Tatbestand,' or else the text is so arranged as to draw out more or less forcibly a complete description of some incident inaccurately mentioned in the narrative.

"Apparatus is arranged so that whenever the subject writes, an electric circuit is closed. It is opened the moment the pencil leaves the paper. A kymograph is used to record these 'makes' and 'breaks' of current. Long, significant pauses may thus be recorded and the correlations between length of pause and character of the supplementation noted. Attempts at dissimulation are easily discovered, first, by a too frequent response of 'nothing'; second, by too many unmeaning reactions; third, by a greatly retarded reaction procedure. At the end of the experiment, Lipmann and Wertheimer had their subject give an introspective account of his mental activity during the examination. Many of those who were found to have betrayed themselves were most positive in their insistence that they had said nothing which could in any way reveal the existence of the complex. Rittershaus sees no value in this method.

"In all of the three above procedures, it is admitted that innocent as well as guilty individuals will make errors in their reproductions. The point to be emphasized, however, is that the guilty make significant ones.

VI. Perception Method (Wertheimer and Klein)

"Optic stimuli are presented by means of a tachistoscope. The exposure is made for a very short time. The association series includes (a) purely irrelevant words, (b) complex-words, (c) irrelevant words similar by sight or sound to those of a complex type, (d) incomplete words which might be interpreted by the subject either as irrelevant or complex, e. g. CH--E may be in-

terpreted in terms of the complex as CHOKe, or may be understood by the innocent individual as CHORE or CHOSE.

"Results seem to indicate that (a) so-called guilty individuals perceive complex stimuli more quickly than irrelevant, (b) irrelevant stimuli are falsely perceived as complex, (c) incomplete stimuli are interpreted more easily as complex than irrelevant.

"Instead of perception being visual, it may be auditory. The authors suggest Gutzmann's procedure in the use of a phonograph. Thus the word *coarse* will be heard as *corpse* by the murderer. Some of the advantages in using a phonograph have been already indicated. Besides the ones mentioned, the phonograph adds obscurity and indistinctness to the stimulus, and eliminates any possibility of lip-reading.

VII. Distraction Method (Wertheimer and Klein)

"The subject is given two texts, the content of one being related to the 'Tatbestand,' while the content of the other is quite irrelevant. He is requested to put a line through all *r*'s as rapidly as possible. The number of *r*'s crossed, the time, errors, and behavior are noted. The results seem to show that the meaning of the text relating to the 'Tatbestand' is much more distracting to the rigid and accurate performance of the task.

VIII. Sentence Method (Moravcsik)

"The experimental material consists largely of sentences expressive of different emotional states. Sentences of indifferent affective toning are also employed. The subject is requested to respond with whatever comes into his mind. No attempt is made to narrow down the reaction to any particular form or type, or to any number of words. Moravcsik experimented only on melancholiacs and maniacs. The following are samples of his material:

Stimulus Words (depression) — *sadness, pain, suffering, grave, sins.*

(exaltation) — *joy, riches, dance, laughter.*

(indifferent) — *garden, house, snow.*

Sentences (depression) — *How unfortunate I am!*

It would be well if I were buried.

My heart is so full of pain.

(exaltation) — *The best thing is a long life.*

I am full of joy!

(indifferent) — *In winter the snow falls.*

He observed that the melancholiac found a group of words expressive of dejection harmonious and desirable, and revolting to those of an exalted type. The reverse was true of the maniac.

394 Diagnosing Personality and Conduct

He also concludes that those possessing a stronger mental balance answer with a single word rather than with a sentence.

IX. The 'Ausfrage' Method (Marbe, Messer, Buhler and others; designated 'Ausfrage' by Wundt)

"The procedure is that of the simple association experiment. A stimulus word is presented, and the subject responds with the first idea coming to consciousness. Immediately following this, however, he is asked to give a complete introspection of all his mental operations during the reaction. This method is of value only where trained introspectionists are acting as subjects. Wundt opposed its use as a psychological experiment, since it lacks the four necessary elements, but Kakise, on the contrary, makes a strong plea for its employment. From one point of view this mode of procedure resembles greatly the usual psychoanalytic method.

"Since some of these methods have only in one or two cases been given a fair and accurate trial by unbiased yet expert psychometrists, for the detection of crime, it is not surprising to find so large a number of workers in this field skeptical as to their value. We certainly need more data. But, granting their worthlessness for criminal psychology, they still are of great promise to experimentalists interested in the diagnosis of individual complexes and constellations. As yet, the procedures are crude and the technique undeveloped. The methods are waiting for a master-hand at reconstruction and improvement."

Summary

That the association method is a powerful tool in the diagnosis of conduct has been demonstrated without a doubt. The earlier attempts at classification have been largely abortive. But the use of *reaction time*, the *reproduction experiment*, certain *repetitions*, and strange *twists* and *misinterpretations* of words are extremely diagnostic of centers of emotional tension. Though the method has been proved to be effective in the discovery of *guilt*, certain difficulties in its use with sophisticated subjects has prevented its practical application. The association method also is helpful in the diagnosis of insanity and milder psychopathic states, although it cannot be used as the sole criterion. Certain interesting and suggestive variations of the free association method have been summarized by Kohs.

REFERENCES

The Free Association Method

1. ALLEN, E. A., "Temperamental Tests," *British Journal of Psychology* (Medical Section), 7: 391-446 (1927).
2. ALLPORT, F. H., *Social Psychology* (Houghton Mifflin Company, 1924).
3. ARNOLD, F., *The Psychology of Association* (Science Press, 1906).
4. BRIDGES, J. W., *An Experimental Study of Decision Types and Their Mental Correlates*, Psychological Monographs, Vol. XVII, 1, No. 72: 51-55 (1914).
5. BRILL, A. A., *Psychoanalysis; Its Theories and Practical Implications* (W. B. Saunders Company, 1913).
6. CATTELL, J. M., and BRYANT, S., "Mental Association Investigated by Experiment," *Mind*, 14: 230-250 (1889).
7. CHASE, H. W., "Psychoanalysis and the Unconscious," *Pedagogical Seminary*, 17: 281-327 (1910).
8. CORIAT, I. H., *Abnormal Psychology* (Dodd, Mead & Company, 1910).
9. CRANE, H. W., *A Study in Association, Reaction and Reaction Time*, Psychological Monographs, Vol. XVIII, 4, No. 80 (1915).
10. DOOLEY, L., "A Study in Correlation of Normal Complexes by Means of the Association Method," *American Journal of Psychology*, 27: 119-151 (1916).
11. DUNLAP, K., "The Stop-Watch and Association Test," *Psychobiology*, 1: 171-175 (1917).
12. EASTMAN, F. C., and ROSANOFF, A. J., "Association in Feeble-minded and Delinquent Children," *American Journal of Insanity*, 69: 125-142 (July, 1912).
13. EKDAHL, A. G., *The Effect of Attitude on Free Word Association-Time*, Genetic Psychology Monographs, Vol. V, No. 4 (April, 1929).
14. FÉRÉ, C., *The Pathology of the Emotions* (London, University Press, 1899).
15. FREUD, M., *The Personalities of the Socially and Mechanically Inclined*. Psychological Monographs, Vol. XXXIII, 4, No. 151 (1924).
16. GALTON, F., *Inquiries into Human Faculty and Its Development* (The Macmillan Company, 1885), pp. 182-203.
17. ———, "Psychometric Experiments," *Brain*, 2: 149-162 (1879); also *19th Century*, 5: 425-533 (1879).
18. GODDARD, H. H., *Juvenile Delinquency* (Dodd, Mead & Company, 1921).

19. GOLDSTEIN, E. R., "Reaction Times and the Consciousness of Deception," *American Journal of Psychology*, 34: 562-581 (1923).
20. GRIFFITHS, C. H., "Results of Some Experiments on Affection Distribution of Associations and Recall," *Journal of Experimental Psychology*, 3: 447-464 (1920).
21. GUTHRIE, E. R., "Measuring Introversion and Extroversion," *Journal of Abnormal and Social Psychology*, 22: 82-88 (1927).
22. HAGGERTY, M. E., and KEMPF, E. J., "Suppression and Substitution as a Factor in Sex Differences," *American Journal of Psychology*, 24: 414-425 (1913).
23. HART, B., *The Psychology of Insanity* (G. P. Putnam's Sons, 1912).
24. HENKE, F. G., and EDDY, M. W., "Mental Diagnosis by the Association Reaction Method," *Psychological Review*, 16: 399-409 (1909).
25. HOCH, A., "Review of Three Books by Jung," *Journal of Abnormal Psychology*, 1: 95-100 (1907).
26. HOLLINGWORTH, H. L., "The Oblivescence of the Disagreeable," *Journal of Philosophy, Psychology and Scientific Method*, 7: 709-714 (1910).
27. HUBBARD, L. M., "Complex Signs in Diagnostic Free Association," *Journal of Experimental Psychology*, 7: 342-357 (1924).
28. HULL, C. L., and LUGOFF, L. S., "Complex Signs in Diagnostic Free Association," *Journal of Experimental Psychology*, 4: 111-136 (1921).
29. JONES, E., *Papers on Psychoanalysis* (London, Baillière, 1913; New York, William Wood & Company, 1923).
30. JUNG, C. G., "The Association Method," *American Journal of Psychology*, 21: 219-269 (1910).
31. ———, "Diagnostische Assoziationsstudien," *J. f. Psychol. u. Neur.*, 7: 1-24, 8: 25-60 (1906); 9: 188-197 (1907).
32. ———, "Diagnostische Assoziationsstudien," *Beiträge zur experimentellen Psychopathologie*, Bd. I (1906); Bd. II (1911); Leipzig, Barth.
33. ———, "On Psychophysical Relations of the Association Experiment," *Journal of Abnormal Psychology*, 1: 247-255, 1907.
34. ———, *Studies in Word Association* (Dodd, Mead & Company, 1918).
35. KATZEN-ELLENBOGEN, E. W., "The Detection of a Case of Simulation of Insanity by Means of the Association Tests," *Journal of Abnormal Psychology*, 6: 19-32 (1911-1912).
36. KELLEY, T. L., "The Association Experiment—Individual Differences and Correlations," *Psychological Review*, 20: 479-504 (1913).

37. KENT, G. H., and ROSANOFF, A. J., "A Study of Association in Insanity," *American Journal of Insanity*, 67: 37-96, 317-390 (1910).
38. KOHS, C. H., "The Association Method in Its Relation to the Complex and Complex Indicators," *American Journal of Psychology*, 25: 544-594 (1914).
39. LANDIS, C., GULLETTE, R., and JACOBSEN, C., "Criteria of Emotionality," *Pedagogical Seminary and Journal of Genetic Psychology*, 32: 209-234 (1925).
40. LANGFELD, H. S., "Psychophysical Symptoms of Deception," *Journal of Abnormal Psychology*, 15: 319-325 (1920).
41. ———, "Suppression with Negative Instruction," *Psychological Review*, 18: 411-424 (1911).
42. LANGFELD, H. S., and ALLPORT, F. H., *An Elementary Laboratory Course in Psychology* (Houghton Mifflin Company, 1916), 147 pp.
43. LASLETT, H. R., "Preliminary Notes on a Test of Delinquent Tendencies," *Journal of Delinquency*, 9: 222-230 (1925).
44. LEACH, H. M., and WASHBURN, M. T., "Some Tests by the Association Reaction Method of Mental Diagnosis," *American Journal of Psychology*, 21: 162-167 (1910).
45. LEWIN, K., "Das Problem der Willinmessung und das Grundgester der Association," *Psychol. Forsch.*, 1: 191-302 (1922).
46. MARSTON, W. M., "Reaction Time Symptoms of Deception," *Journal of Experimental Psychology*, 3: 72-87 (1920).
47. MATEER, F., "The Future of Clinical Psychology," *Journal of Delinquency*, 6: 283-293 (1911).
48. McCLATCHY, V. R., "Differences in the Oral Responses to Words of General and of Local Significance," *Journal of Experimental Psychology*, 5: 312-322 (1922).
49. MEYER, A., "Normal and Abnormal Associations," *Psychological Bulletin*, 2: 242-259 (1905).
50. MITCHELL, I., ROSANOFF, I. R., and ROSANOFF, A. J., "A Study of Association in Negro Children," *Psychological Review*, 26: 354-359 (1919).
51. MOORE, H. T., "A Method of Testing the Strength of the Instincts," *American Journal of Psychology*, 27: 227-233 (1916).
52. MOORE, H. T., and GILLILAND, A. R., "The Measurement of Aggressiveness," *Journal of Applied Psychology*, 5: 97-118 (1921).
53. MORGAN, E., MULL, H. K., and WASHBURN, M. F., "An Attempt to Test Moods or Temperaments of Cheerfulness and Depression by Directed Recall of Emotionally Toned Experiences," *American Journal of Psychology*, 30: 302-304 (1919).

398 Diagnosing Personality and Conduct

54. MÜNSTERBERG, H., *On the Witness Stand* (The McClure Co., 1908), pp. 73-110.
55. ———, "The Third Degree," *McClure's Magazine*, 29: 614-622 (1907).
56. MURPHY, G., "An Experimental Study of Literary versus Scientific Types," *American Journal of Psychology*, 28: 238-263 (1917).
57. O'CONNOR, J., *Born That Way* (The Williams and Wilkins Company, 1928).
58. PETERSON, F., and JUNG, C. G., "Psychophysical Investigations with the Galvanometer and Pneumograph in Normal and Insane Individuals," *Brain*, 30: 153-218 (1907).
59. PRINCE, M., *The Unconscious* (The Macmillan Company, 1914).
60. REINHOLD, F., "Beiträge zur Assoziationslehre auf Grund von Massenversuchen," *Zeitsch. f. Psychol.*, 54: 183-214 (1909).
61. ROSANOFF, A. J., *Manual of Psychiatry* (John Wiley and Sons, Chicago, 1905-1927), ch. VII.
62. ROSANOFF, A. J., MARTIN, H. E., and ROSANOFF, I. R., *A Higher Scale of Mental Measurement and Its Application to Cases of Insanity*, Psychological Monographs, Vol. XXV, 3, No. 109 (1918).
63. ROSANOFF, I. R., and ROSANOFF, A. J., "A Study of Association in Children," *Psychological Review*, 20: 43-89 (1913).
64. RUSK, R. R., "Experiments on Mental Association in School Children," *British Journal of Psychology*, 3: 349-355 (1909-1910).
65. SALLING, G., "Assoziative Massenversuche," *Zeitsch. f. Psy.*, 49: 238-253 (1908).
66. SMITH, W. W., "Experiments on the Association Tests as a Criterion of Individuality," *British Journal of Psychology* (Medical Section), 2: 121-130 (1922).
67. ———, "Experiments on Memory and Affective Tone," *British Journal of Psychology*, 11: 236-250 (1921).
68. ———, *The Measurement of Emotion* (Harcourt, Brace and Company, 1922).
69. ———, "The Relation between Complex Indicators and the Form of the Association," *British Journal of Psychology* (Medical Section), 1: 297-315 (1921).
70. ———, "Some Properties of Complex Indicators," *British Journal of Psychology* (Medical Section), 1: 281-296 (1921).
71. STRONG, E. K., JR., "A Comparison between Experimental Data and Clinical Results in Manic-Depressive Insanity," *American Journal of Psychology*, 24: 66-98 (Jan., 1913).
72. STUMBERG, D., "A Comparison of Sophisticated and Naïve

- Subjects by the Association-Reaction Method," *American Journal of Psychology*, 36: 88-95 (1925).
73. TERMAN, L. M., "Adventures in Stupidity: A Partial Analysis of the Intellectual Inferiority of a College Student," *Science Monthly*, 14: 24-40 (1922).
 74. ———, *Genetic Studies of Genius*, Vol. I, 2d ed. (Stanford University Press, 1925).
 75. TOLMAN, E. C., and JOHNSON, I., "A Note on Reaction Time and Feeling," *American Journal of Psychology*, 29: 187-195 (1918).
 76. WASHBURN, M. F., and others, "Directed Egocentric Reactions," *American Journal of Psychology*, 30: 300-302 (1919).
 77. WATSON, J. B., *Psychology from the Standpoint of a Behaviorist* (J. B. Lippincott Company, 1919).
 78. WELLS, F. L., "Association Types and Personality," *Psychological Review*, 26: 371-381 (1919).
 79. ———, *Mental Adjustments* (D. Appleton and Company, 1917).
 80. ———, "The Personal Factor in Association Reaction," *American Journal of Insanity*, 69: 897-906 (1913).
 81. ———, "Practice Effects in Free Association," *American Journal of Psychology*, 22: 1-13 (1911).
 82. ———, "A Preliminary Note on the Categories of Association Reactions," *Psychological Review*, 18: 229-233 (1911).
 83. ———, "The Question of Association Types," *Psychological Review*, 19: 253-270 (1912).
 84. ———, "Some Properties of the Free Association Time," *Psychological Review*, 18: 1-23 (1911).
 85. WELLS, F. L., and FORBES, A., "On Certain Electrical Processes in the Human Body and Their Relation to Emotional Reactions," *Archives of Psychology*, Vol. II, No. 16 (1911).
 86. WELLS, F. L., and ROONEY, J. S., "A Simple Voice Key," *Journal of Experimental Psychology*, 5: 419-427 (1922).
 87. WOODWORTH, R. S., and WELLS, F. L., *The Association Test*, Psychological Monographs, Vol. XXII, 5: No. 57 (1911).
 88. WOODROW, H., and LOWELL, F., *Children's Association Frequency Tables*, Psychological Monographs, Vol. XXII, 5: No. 97 (1916).
 89. WRESCHNER, A., "Eine experimentelle Studie über die Assoziation in einem Falle von Idiotie," *Allgemeine Zeitschrift für Psychiatrie*, 57: 241-339 (1900).
 90. WUNDT, W., *Grundzüge der Physiologischen Psychologie* (Leipzig, Engelmann, 1911), III, 436-456, 519-543.
 91. YERKES, R. M., and BERRY, C. S., "The Association Reaction Method of Mental Diagnosis," *American Journal of Psychology*, 20: 22-37 (1909).

Chapter XI

PHYSIOLOGICAL MEASURES OF THE EMOTIONS

CONDUCT which is accompanied by emotion has special significance. In certain emotions the body is prepared for immediate and intense muscular activity, through the functioning of the sympathetic nervous system. All the forces of the body are adjusted to permit vigorous combat or flight. The heart pumps more vigorously to carry away the fatigue poisons, and breathing becomes deeper and faster to aid in carrying away the waste carbon dioxide products resulting from increased metabolism. Blood is withdrawn from the digestive organs so that the muscles may have a greater supply, and the gastric secretions are retarded. The pupils of the eye enlarge to permit the entrance of more light. The sweat-glands secrete, thus acting as a temperature regulator against the increased heat to be produced. The liver releases sugar into the blood for an enlargement of immediately available energy, and adrenalin is secreted to facilitate the removal of fatigue products and to hasten blood-clotting if necessary.

A man in a baffling situation has various alternatives before him. He may postpone immediate action in favor of deliberation. After a careful survey of the possibilities for resolving the conflict, he may make a decision and then calmly and coolly plan to put his decision into effect. At the other extreme he may make an immediate and vigorous attempt to free himself from the baffling situation which confronts him according to his past habit systems. In this case his impulsive behavior receives reinforcement from the sympathetic nervous system, and we may say that he is reacting *emotionally*. If this is the method employed, there is a minimum of thought used. Usually one meets life's difficulties by a combination of deliberation and emotion.

It must be evident from the above description that the "emergency" type of behavior has only a fair chance of being satisfactory. If previous habit has provided a good adjustment to the

novel and irritating situation in the past, then the response may be satisfactory. But this method is blind. Only when the various alternatives in the situation have been fairly surveyed can it be said that the response is reasonable or intelligent. Emotional conduct is notoriously blind and haphazard.

Because conduct accompanied by emotion has peculiar significance, it is desirable that we have measures of emotional activity. One method of measuring this emotional activity is by testing for conduct which results in poor adjustment. The Woodworth Psychoneurotic Inventory has been called a test of emotional stability. As a test of the emotions this implies that all of the conduct, the presence of which is indicated on the questionnaire, is accompanied by emotion. This is on the whole probably a fair assumption, but the relation between the conduct and the presence of emotion may not be very direct. What is needed is a more direct measure of the visceral changes which are characteristic of the emotional, stirred-up state.

Of the visceral changes which characterize the emotions, some are inaccessible or inconvenient to get at. Rubber balloons to be swallowed by the subject have been employed to measure changes in the peristaltic movements of the stomach, duodenum, and rectum by Brunswick,* but the method is so distasteful to the subjects that its use is very restricted. Likewise, detection of the presence of sugar or adrenin in the blood requires high technical skill. The changes which are most amenable to accurate measurement are (a) the rate of the heart-beat, (b) the amplitude of the heart-beat, (c) blood-pressure, (d) volume of blood in a limb, (e) rate of breathing, (f) amplitude of breathing, and (g) psychogalvanic reflex.

Before these methods are described in detail a few general words concerning laboratory techniques are in order. The methods of testing for the presence of emotion require elaborate and delicate instruments which must be carefully set up in the laboratory. Consequently the measurements must be carried out in the laboratory usually under somewhat artificial conditions. Situations whose effect on conduct are to be tested must be imported into the laboratory. Since blood-pressure, pulse rate, breathing, and perspiration vary irregularly and seem to be influenced by

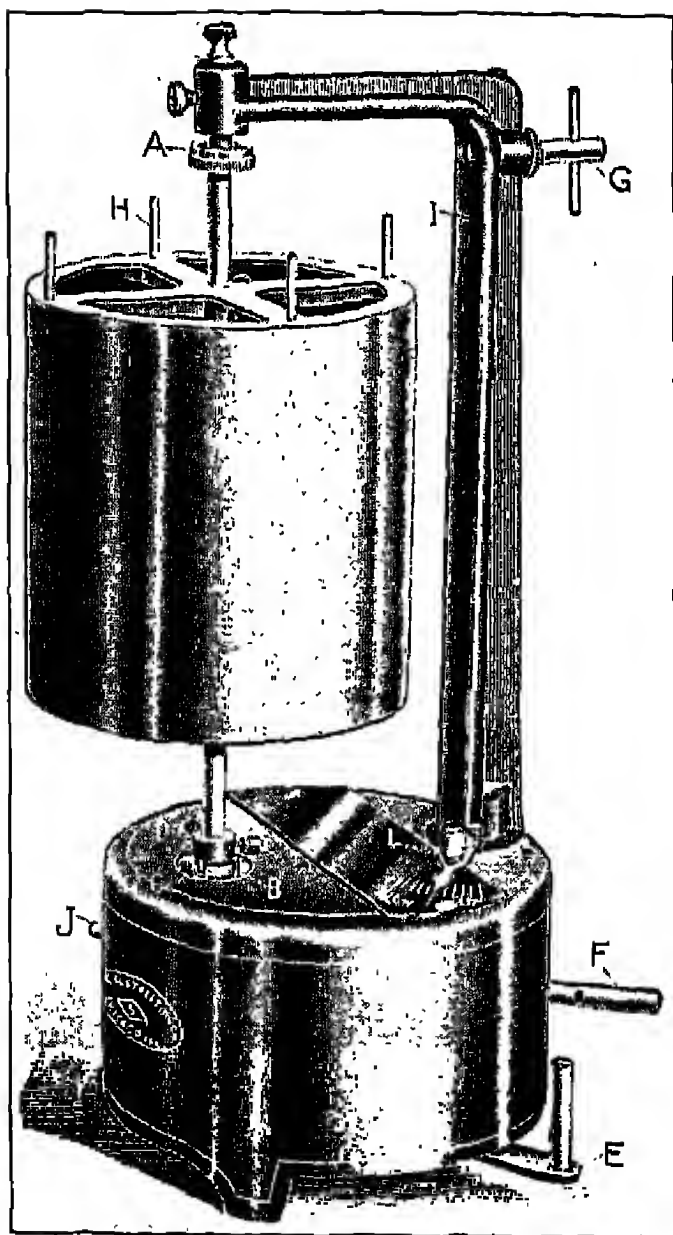
* Brunswick, D., "The Effects of Emotional Stimuli on the Gastro-Intestinal Tone," *Journal of Comparative Psychology*, 4: 19-80, 225-288 (1924).

a large number of obscure factors, all changes within a single individual must be noted during the experiment itself. In paper-and-pencil testing it is possible to compare one individual with another, one group with another, or the response of an individual at one time with his response at another time. But with these physiological measures it is not possible to compare one individual with another or one group with another or an individual's record at one sitting with his record at another sitting. All comparisons must be made between changes in the record during a continuous sitting when the environment is under control and relatively constant. Even so, metabolic processes within the body or the influence of ideas produce changes which are impossible to control or interpret.

In ordinary clinical practice all observations are noted by the experimenter and recorded by him. All observations represent a cross-section picture of a continuous process. Pulse is counted for a fraction of a minute and recorded. Changes in pulse rate can be obtained by taking separate pulse counts and noting the change in rate. Or blood-pressure is taken at a given moment and recorded. Changes in blood-pressure are determined by taking two or more separate readings and subtracting to determine the amount of change.

For accurate and precise work a continuous record is very desirable. In the first place, if the record can be recorded mechanically, one possible cause of error which occurs in observation and the recording of observation is eliminated. In the second place, a continuous record permits much more precise information as to the time when changes occur, and the amount of these changes. Smaller movements or changes that may have some significance will utterly escape the experimenter who must measure the process at infrequent intervals, but these can be accurately noted on a continuous record.

The instrument used in physiology for recording such continuous changes is known as the *kymograph* (curve-writer), invented by the famous physiologist Ludwig. This consists of a metal cylinder (A), called the drum, which may be made to revolve slowly at a constant and determined rate by means of a clock-work (B). Around this metal cylinder is wrapped and glued a sheet of glazed paper. This paper is smoked by moving it swiftly over a gas flame until it is coated with a thin, uniform layer



KYMOGRAPH (curve-writer)

of soot. This smoked paper surface is then used as a recording surface. The record is made by the tracing of the point of a lever arm (tambour),* which is moved by the phenomenon (pulse or breathing) being studied. The point of the lever arm is brought lightly in contact with the smoked paper, and the record is made by scratching a path through the sooty surface. This tambour or lever arm is placed at right angles to the axis of the cylinder so that ordinates of the kymograph record represent the amplitude of the phenomenon being investigated, while the abscissæ are time intervals. Several tambours may be in operation at the same time so as to give a simultaneous record on the same kymograph sheet. Since the revolving mechanism does not usually record the speed of revolution it is necessary to have a special time marker or tambour which is made to move at regular intervals of a second or less by clock-work. When the record is completed, the paper is slipped through a bath of diluted white shellac which coats the record and makes it permanent.

In more recent developments sensitive photographic paper has been used on the kymograph cylinder. This is made to revolve behind a narrow slit. A moving spot of light may be made to focus on this slit, thereby giving a record. Or the lever arm may be placed in the path of a bright light and made to send its focused shadow on the slit where the shadow is recorded on the sensitive paper. This photographic method is used in recording the movements of the string galvanometer.

Measurement of the Rate of the Heart-Beat—the Pulse

The pulse may be measured simply by pressing the radial artery in the wrist and counting the number of beats for a given

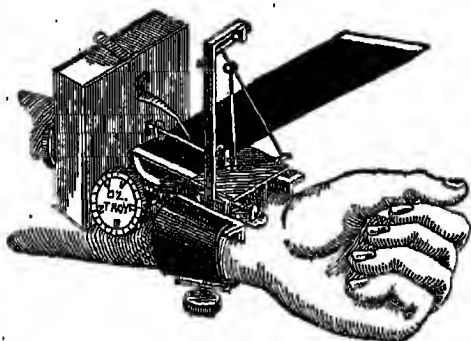
*The more simple tambours work by means of a system of levers. Sometimes, however, it is not convenient to transmit movements to the kymograph by means of levers. In such cases, when the kymograph is some distance from the instrument, a device known as "Marey's tambour" is used. This consists of a metal tray or basin covered with a rubber membrane. This rubber membrane may be made to move up and down with certain pressure changes in the rubber tube which connects it with the measuring instrument. The lever arm is connected directly with this rubber membrane and records its movements.

Sometimes two Marey's tambours are employed, one of which is on the measuring instrument itself, the membrane of which is stimulated directly. A rubber tube conveys changes in pressure in this first Marey's tambour to the second at the kymograph some distance away. The two rubber membranes must oscillate in unison.

404 Diagnosing Personality and Conduct

number of seconds. An instrument for recording the pulse-beat is known as the *sphygmograph*, which not only will give the pulse rate but furnishes a detailed picture of the character of the arterial pulse. This latter, however, is not needed in studying the emotions.

The first *sphygmograph* was invented by Marey, and is relatively simple. A *pelotte* or button is placed directly over the radial artery in the wrist. The movements transmitted to it are relatively small. These must be magnified by a series of levers until



THE DUDGEON SPHYGMOGRAPH IN POSITION

From Howell's *Text-Book of Physiology*, W. B. Saunders Company, 1924.

finally the movement is transmitted to the writing lever arm which makes the record directly on the kymograph.

The Dudgeon sphygmograph is easier to use than the Marey instrument. It has two improvements. One is a device by means of which the pressure of the button on the artery can be graduated. The other and most important variation is that the kymograph is an essential part of the instrument and consequently the arrangement of levers can be better controlled. Even in the best of these instruments, however, there is a certain amount of looseness in the lever joints giving an inertia which prevents the graphic picture from corresponding exactly with the pulse wave. However, it is sufficiently accurate to afford a measure of pulse rate.

Pulse rate has not been used extensively as a measure of emotional excitement, even though it is easy of determination, mainly because it is subject to such a variety of forces. The important

Physiological Measures of the Emotions 405

factors which influence pulse are given below, the first three mentioned representing rather fixed adjustments.

1. *Variations with sex.* The pulse rate of women is known to be higher than that of men at all periods of life.

2. *Variations with size.* The larger an individual the slower his pulse, other things being equal. This may in part explain the sex difference. Large animals have very definitely slower pulse rates than small animals, and some small birds are said to have pulse rates of several hundred beats a minute.

3. *Variations with age.* As one gets older, the pulse rate grows less. In extreme old age, however, there is a slight acceleration, due perhaps to selection or to shrinkage of body volume. The following table * gives the approximate average rates.

TABLE 79
PULSE RATES AT VARIOUS AGES

	<i>Pulse beats per minutes</i>
At birth	140
Infancy	120
Childhood	100
Youth	90
Adult age	75
Old age	70
Extreme old age	75-80

4. *Variations with temperature.* The pulse rate varies with the temperature of the blood. As there is an increase in the temperature of the blood there is a corresponding increase in heart-beat. In cold-blooded animals the heart-beat is very sluggish.

5. *Variations with the presence of drugs in the blood stream.* Certain drugs, notably adrenalin, will increase the rate of the heart-beat directly. Other drugs, such as potassium salts, will retard or even entirely stop the heart action.

6. *Variations with exercise.* The next five factors are thought to influence pulse rate through the action of the nervous system. First, it is well known that muscular exercise markedly affects the rate of the heart-beat. This effect, while not ordinarily noticed, may be produced by even very light muscular activity such as tapping a telegraph key. It is probably this factor more

* See Burton-Oritz, R., *A Textbook of Physiology* (W. B. Saunders Company, 1920), p. 273.

than any other that makes it so difficult to use the changes in pulse rate as a measure of emotions. This increased pulse rate is an adaptive mechanism for providing the muscular system with an increased supply of blood.

7. *Variations in posture.* The pulse rate is higher when standing than when sitting, and higher when sitting than when lying down. This phenomenon may be included under the foregoing one, inasmuch as the difference is probably due to differences in tone of the postural muscles.

8. *Variations with eating.* The pulse rate is increased after food is received into the stomach. This is believed to be a regulating mechanism to counteract the effect on blood-pressure of the large vascular dilatation in the intestinal area.

9. *Variations with swallowing.* Starling* states that the act of swallowing causes a reflex quickening of the rate of the heart-beat by inhibition of the tonic vagus action.

10. *Variations with blood-pressure.* There is an inverse relation between blood-pressure and heart rate. Low blood-pressure tends to quicken heart rate, while high blood-pressure tends to make the heart beat more slowly. This is an adaptive mechanism for equalizing the blood supply to the tissues under differences in blood-pressure.

11. *Variations due to emotion.* Besides all of these factors influencing heart rate, there is left the direct influence of emotion: the direct accelerating or inhibiting action of the nervous system. It is well known that the motor nerve impulses traveling over the vagus nerve tend to slow up the heart-beat, whereas the impulses passing out through the sympathetic nervous system tend to accelerate the heart-beat. These two sets of controls work in a truly antagonistic manner at all times, and thereby exert a regulating mechanism on the heart rate. Sometimes the sympathetic system is aroused directly by peripheral stimulation, thereby causing an increase in the rate of the heart-beat. But there is also an interrelationship between the various factors. Exercise, blood-pressure, and presence of adrenalin in the blood stream are also associated with emotional excitement so that there is an interrelationship between the various factors controlling the pulse rate.

* Starling, E. H., *Principles of Human Physiology* (London, J. & A. Churchill), p. 803.

Because the pulse is so sensitive to a number of factors, some of which are related to the emotions only remotely and some not at all, pulse rate is difficult to use as a measure of the emotions.

Measurement of the Amplitude of the Heart-Beat

Besides the rate of the heart-beat, the *amplitude* of the heart-beat is known to vary with emotional excitement. Usually *rate* and *amplitude* go together. As one increases, the other increases also. But they do not always run in parallel this way. Physiologists believe that different fibers from the vagus and also from the sympathetic system control rate and amplitude of the heart-beat.

The sphygmograph, which gives a graphic record of the pulse wave, may be used to give a measure of the amplitude of the heart-beat as well as of the rate.

The factors which govern rate also in general govern amplitude. Under just what conditions one is accelerated or inhibited more than the other has been studied in some detail by Eng (4). Her findings are as follows:

	<i>Pulse height</i>	<i>Pulse rate</i>
Attention	Decrease	Retarded
Physical work	Increase	Accelerated
Displeasing taste and smell	Decrease	Accelerated
Pain	Decrease	Accelerated
Fright	Increase followed by decrease	Accelerated, then retarded
Displeasure and excitement	High	Accelerated and irregular
Pleasing smell	Increase	Retarded
Pleasing taste	Increase	Accelerated
Pleasure resulting from other sensory impressions	Increase	Retarded

Measurement of Blood Pressure

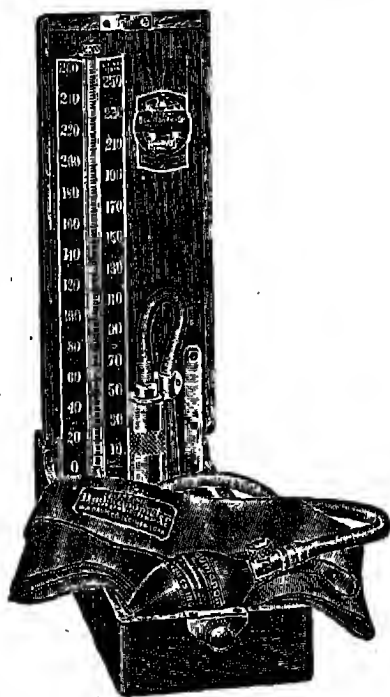
Some investigators, notably Marston and Larson, have made extensive use of blood-pressure in the measurement of emotional excitement and claim to have achieved satisfactory results.

It is well known in physiology that a certain pressure is main-

408 Diagnosing Personality and Conduct

tained against the walls of the arteries due to the force of the heart-beat, the resistance in the capillaries, and the pressure of the arterial walls. This arterial pressure is exceedingly sensitive to a number of factors, among which may be included emotional excitement.

This arterial or blood-pressure is measured by noting the force or pressure necessary to collapse the arterial wall and prevent the pulse beat from passing a given point.



SPHYGMOMANOMETER

The instrument for causing this collapse of the arterial wall and measuring the pressure necessary to do this is known as a *sphygmomanometer*. (Pronounce this in two parts—*sphygmo-manometer*. *Sphygmo-* comes from the Greek meaning "pulse." A *manometer* is an instrument used in physics for measuring pressure. Consequently a *sphygmo-manometer* is an instrument for measuring blood-pressure.) The instrument commonly used is known as the Riva Rocci sphygmomanometer, and is really

quite simple in construction and operation. It consists of three parts. The first part is for compressing an artery. This is usually accomplished by wrapping a band (nowadays usually made of silk) around the arm just above the elbow. Inside this band is a rubber bag which also encircles the arm. This rubber bag is inflated with air and by an increase in the pressure of the air in the rubber bag a corresponding increase of pressure is communicated to the arm. This pressure may be raised to such a point that the brachial artery is made to collapse, thereby preventing the pulse from coming through. Pressure is raised by means of a rubber bulb squeezed by the hand.

The rubber tube by which the air pressure is communicated to the pressure bag on the arm is also connected with a *manometer* or pressure gage. This consists of a U-tube of mercury. One end of the tube is open to the air; the other receives the pressure. As the pressure is increased, it forces the mercury down on one side of the tube and up on the other. The height to which the mercury rises may be read on a graduated scale placed behind the tube, and this reading becomes a measure of the pressure necessary to collapse the artery. Sometimes a calibrated aneroid gage is used in place of the mercury manometer, but this instrument is not so accurate as the mercury manometer.

In applying the instrument the fingers of the left hand are placed on the radial artery of the wrist while the right hand is employed to inflate the rubber pouch. At the moment that the pulse disappears a reading is taken which is known as the *systolic pressure*. Sometimes this is checked by raising the pressure above that necessary to stop the pulse and then noting the pressure reading at the point where the pulse beat first breaks through. In the *auscultatory method* (*auscultation* means "the act of listening," hence the *auscultatory method* is the method whereby impressions are gained by listening), the pulse is observed by using the *binaural stethoscope*. (A *stethoscope* is an instrument used by physicians and others for hearing more distinctly than can be done with the ear alone sounds coming from within the body, particularly the chest. *Stetho-* comes from the Greek meaning "chest." The use of the stethoscope is not, however, confined to the chest. *Binaural* means that the stethoscope conveys the sounds to both ears.) The stethoscope is applied to the brachial artery just below the point where the pressure band

410 Diagnosing Personality and Conduct

of the sphygmomanometer is applied. The pressure is raised above the systolic level and then allowed to fall slowly by means of a needle valve. At the first point when the sound breaks through, the manometer reading should be recorded as the *systolic pressure*. By allowing the pressure to drop still further the diastolic pressure may also be approximately observed. Ettinger, who has studied the method extensively, notes five phases of the sound in the stethoscope as the pressure is allowed to fall.

1. The initial clear, sharp sound—systolic pressure.
2. The sounds become more continuous, muffled, like a murmur.
3. The sounds become progressively clearer and louder.
4. The sounds become muffled and dull.
5. The sounds disappear.

According to later workers, the diastolic pressure is indicated most accurately by the transition from phase 3 to phase 4, at the beginning of phase 4.

Systolic pressure is usually employed in testing for the presence of emotion.

It will be noted that this method of observing blood-pressure does not permit a continuous reading. Indeed, it is only by means of a somewhat complicated apparatus, known as the Erlanger sphygmomanometer, that the readings for a single testing of blood-pressure can be recorded on a kymograph.

Koll, however, has described an instrument for obtaining a continuous reading of blood-pressure. Koll's (15) method and apparatus for giving a continuous record of blood-pressure by the indirect method are not very well known. Yet the principle involved is so simple as to warrant more extensive experimentation with the method.

Koll reports that continuous observations may be made for fifteen or twenty minutes, after which the continued pressure on the arm causes considerable discomfort. Ordinarily in using the instrument one takes a tracing for four or five minutes, deflates the pressure for a minute or so, and then resumes the tracing.

The factors which influence blood-pressure are many indeed. They will be enumerated in order with little comment. It may be seen by following through the various factors that blood-

Physiological Measures of the Emotions 411

pressure is one of the major factors in determining daily or even hourly efficiency. In the following summary the pressure referred to is that in the arteries.

1. *Blood-pressure varies directly with the volume of the blood.* Extensive blood-letting is accompanied by a decided fall in pressure.

2. *Blood-pressure varies directly with the energy of the heart,* other things remaining constant. The pressure increases with an increase in the rate and force of the heart-beat. In accurate measurements of blood-pressure in animals, in which an artery is opened and connected directly with a manometer, it is found that the pressure varies with the cycles of the heart-beat. The pressure is strongest at the time of the ventricular systole.

3. *Blood-pressure varies with the elasticity of the blood-vessels.* This is partly a true elasticity of the tissues constituting the arterial walls and is partly a state of tonus of the smooth musculature which form a part of the wall of every artery.

4. *Blood-pressure varies with the peripheral resistance.* That is, blood-pressure is controlled in part by the resistance of the smaller arteries and capillaries farther along in the arterial system.

5. *Gravity causes variations in blood-pressure.* When one is erect this force tends to increase pressure in arteries at levels below the heart and to decrease it at levels above the heart. Upon lying down a distinct change in the relative pressure is effected.

6. *Blood-pressure is higher on the average in man than in woman.*

7. *Blood-pressure varies with the size of the animal.* In general the larger the animal, the higher the pressure.

8. *Blood-pressure increases with age,* as is shown in the following table taken from Burton-Opitz. In old age blood-pressure increases due to the loss of elasticity of the vascular tissue.

9. *Breathing causes variations in blood-pressure.* As the breath is taken in, suction is exerted upon the blood-vessels in the thorax and the large veins near the heart, causing a fall in pressure. The converse effects are noted when the breath is expelled. The breathing consequently causes small, undulating changes in pressure throughout the arterial system. Deep and forced breathing increases the pressure in general.

TABLE 80
BLOOD-PRESSURES AT VARIOUS AGES
(from Burton-Opitz *)

	Blood-pressure in millimeters of mercury
First few months	70- 75
1- 2 years	80- 90
2- 3 years	90-100
3-10 years	95-115
10-15 years	100-115
15-20 years	105-128
20-30 years	135
30-40 years	140
40-50 years	142
50-60 years	154
60-70 years	180

10. *A rhythmical rise and fall of blood-pressure is known as the Traube-Hering waves.* Each of these waves is longer than those due to respiratory movements. These waves are believed to be due to a rhythmical action of the vasomotor center.

11. *Blood-pressure is higher after food has been taken into the system.* This may be due to a concentration of blood in the abdominal organs, while in the splanchnic area the pressure is lowered.

12. *Muscular exercise causes a rise in blood-pressure.* To this may be coupled other factors. Blood-pressure is lower during sleep, probably largely because of the cessation of muscular activity. Blood-pressure rises during mental work. While this may represent a distinct phenomenon, it probably is due to the increased muscular activity which is concomitant with mental labor.

13. *Cold baths as well as hot baths produce a rise in blood-pressure,* although for a very different reason. In the one case there is a marked constriction of the surface arteries; in the other the increase is due to an increase in the frequency of the heart. Baths at the temperature of the body have no influence on blood-pressure.

14. *Blood-pressure falls with fatigue.* Such substances as lactic acid and carbon dioxide, waste products of muscular activity, cause a relaxation of muscular tonus and produce a dilation of the arteries, thus providing more blood at a time when it is needed.

* Burton-Opitz, R., *A Textbook of Physiology* (W. B. Saunders Company, 1920), p. 370.

On the other hand, it is well known that the presence of carbon dioxide in the blood causes a rise in arterial pressure through the vasomotor centers. Starling* harmonizes these two opposite influences as follows:

"We thus see that carbon dioxide, which is the universal hormone set free in the circulation when the activity of the body as a whole is increased, has a double effect on the blood-vessels—a central effect through the vasomotor centers, medulla, and spinal cord, causing contraction of the blood-vessels, and a local peripheral effect causing dilatation of the blood-vessels. The general result therefore will be to cause dilatation of the blood-vessels of the part where the carbon dioxide is produced and where it is present in greatest concentration, and vascular constriction elsewhere under the influence of the sensitive nervous centers."

15. *Blood-pressure is lowered during menstruation and raised during pregnancy, and is remarkably higher during labor.*

16. *Pain causes a rise in blood-pressure.*

17. *Cold produces a rise in blood-pressure.*

18. *Certain glandular products cause marked changes in blood-pressure.* Adrenalin produces a marked constriction of the arterioles and a rise in blood-pressure. Thyroid deficiency causes a drop in blood-pressure, while thyroid extract, when fed, causes a faster pulse and higher blood-pressure. Likewise pituitary extract causes a rise in blood-pressure.

19. Finally we come to the *effect of emotion on blood-pressure*. In general emotional excitement causes a rise in blood-pressure.

It should be observed that there is an intricate interdependence among these various factors. The introduction of one factor as a stimulus causes a complex readjustment in the body. For instance, under the influence of an emotional stimulus blood-pressure is directly increased. But the heart rate and force also increase, breathing increases in rate and in amplitude, and adrenalin is poured into the blood stream. These reactions also have their own effect on blood-pressure. Blood-pressure in turn has its influence on the other factors. The final outcome is an intricate adjustment and balance of the various reciprocal factors.

It should also be observed that while on the whole the vascular system works as a unit, there are special controls for distributing

* Starling, E. H., *Principles of Human Physiology* (London, J. & A. Churchill), p. 850.

414 Diagnosing Personality and Conduct

the blood supply to different parts of the body. In general there are vaso-constrictor and vaso-dilator fibers which work antagonistically. Concerning this Howell says,

"It may be supposed that under normal conditions the activity of this mechanism is adjusted so as to control the blood-flow through the different organs in proportion to their needs. When the blood-vessels of a given organ are constricted the flow through that organ is diminished, while that through the rest of the body is increased to a greater or less extent corresponding to the size of the area involved in the constriction. When the blood-vessels of a given organ are dilated the blood-flow through that organ is increased and that through the rest of the body diminished more or less. The adaptability of the vascular system is wonderfully complete, is worked out mainly through the reflex activity of the nervous system exerted partly through the vasomotor fibers and partly through the regulatory nerves of the heart."*

It must be evident from all of the foregoing that it is difficult to use blood-pressure as a simple indication of the pressure of emotion. So many factors operate to influence blood-pressure that conditions must be very carefully controlled if the presence of emotion is to be detected. One thing is certain—when this is our purpose, we cannot compare blood-pressure obtained at one sitting with pressure at another sitting, or pressure in one person with that in another person. In experimental work all comparisons must be made between changes in the same individual during the same sitting, so that all factors may be under rigid control. Even at best there is certain to be variation in the forces acting, due perhaps to metabolic processes going on within the body or to the emotional influence of stray thoughts. Experimenters report just such unexplainable variations. If blood-pressure is to be used in the study of emotion, it must probably be used when the emotion is so very intense as to outweigh in effect any of the lesser variable forces acting.

Measurement of Blood Volume

Still another index of the changes in the circulatory system is a measure of the blood supply in an organ, as determined by measuring the volume of some part of the body, such as the

*Howell, W. H., *A Textbook of Physiology* (W. B. Saunders Company, 1929), p. 622.

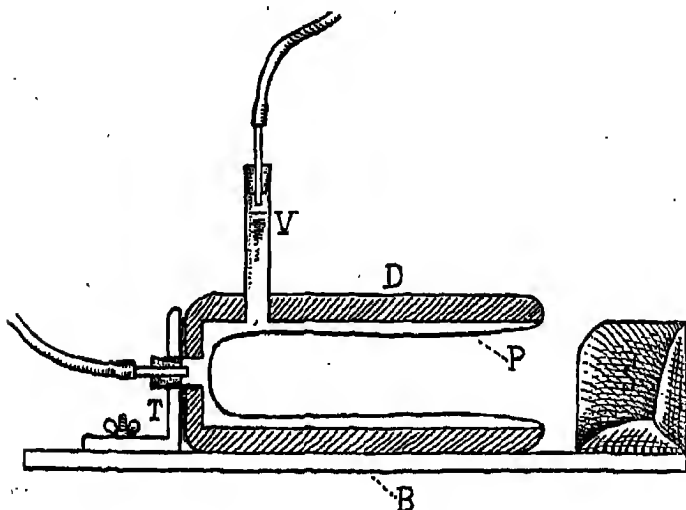
hand and forearm. The measurement of volume is complementary to measurement of pressure and pulse. As the arteries and arterioles in an organ become constricted, with consequent diminution in the volume of the organ, the pressure on the arteries supplying blood to that organ becomes higher. Conversely, as the volume becomes greater, the pressure becomes less. Also, as the pulse increases in rate and amplitude, the volume in the peripheral musculature tends to increase, and vice versa. Since emotion has an influence on the vasomotor system, the measurement of the volume of an organ is another important means of estimating the presence and amount of emotion.

The instrument for measuring blood volume is called a *plethysmograph* (*plethysmo-* from the Greek meaning "enlargement"). The plethysmograph in reality is an instrument for measuring the volume of a limb, but because changes in the volume of a limb are due to changes in blood volume, the instrument becomes an instrument for measuring blood volume. The most common form of the plethysmograph is that used by Lehmann, which measures the volume of the forearm and hand. The following description is borrowed from Eng (4, pp. 5-7), who made practical use of the Lehmann plethysmograph in her experimental work.

"The apparatus employed to record the volume-pulse curve of the arm was an improved form of plethysmograph devised by Lehmann. It consists of a metallic cylinder open at one end. Inside the metal cylinder lies a sleeve (P) of the finest soft rubber which is attached to the rim of the cylinder, forming a water-tight union. During the experiments the arm rests in the rubber sleeve; the space between the sleeve and the cylinder is filled with water, so that the sleeve is moulded to the hand by the pressure of the water and covers it like a glove. The water is admitted through a small opening at the farther end of the cylinder which is connected by means of a short pipe and a piece of rubber tubing to a glass funnel fixed at the requisite height on a stand. A short pipe from the side of the cylinder opens at the top into a tube 10-14 cm. long and of about 1 cm. in diameter, which acts as a water manometer (V). The tube is closed above by a rubber stopper through which a piece of glass tubing passes, and the glass tubing is connected by means of rubber tubing to the recording apparatus.

"The water flows into the plethysmograph until it reaches up into the glass tubing; every further expansion of the arm (whether it be caused by individual pulse beats or by more marked

changes) drives the water somewhat higher up, and is thus made visible by the change in the surface-level of the water. If the volume decreases, the water-level sinks correspondingly. The changes of pressure thus produced are transmitted through the air in the rubber tubing to the membrane of the recording apparatus, and are inscribed on the kymograph. The necessary



PLETHYSMOGRAPH

From Eng, Helga, *Experimental Investigations Into the Emotional Life of the Child Compared with That of the Adult* (London, Humphrey Milford, Oxford University Press, 1925; New York, Oxford University Press), p. 6.

pressure on the arm is adjusted according to the readings of the water manometer; it must not be too weak or the rubber sleeve will not envelop the arm closely enough; nor must it be too strong, as in that case it would give rise to an ill-defined curve. The correct amount of pressure for the best curve must therefore be decided by trial. The pressure is regulated by raising or lowering the funnel, which is therefore called the manometer funnel."

Eng describes certain refinements in the method based on her experience, and her report should be consulted by anyone planning to experiment with the apparatus.

The factors which control blood volume are similar to those controlling pulse rate and height and blood-pressure, since volume is a resultant of these other phenomena working in combination.

Physiological Measures of the Emotions 417

Eng notes the following changes in volume due to various forms of activity and emotional excitement.

Attention-tension	Falling
Psychical work	Rising
Displeasing taste or smell	Falling
Pain	Falling
Fright	Rising followed by falling
Spontaneous displeasure due to sensory impressions	Marked falling
Psychical displeasure (unpleasant concepts)	Low with abrupt rising and slower fallings
Displeasure and excitement	High
Depression	Low and even
Pleasing smell	Rising
Pleasing taste	Rising
Spontaneous pleasure from sensory impressions or concepts	Rising

Measurement of Rate and Amplitude of Breathing

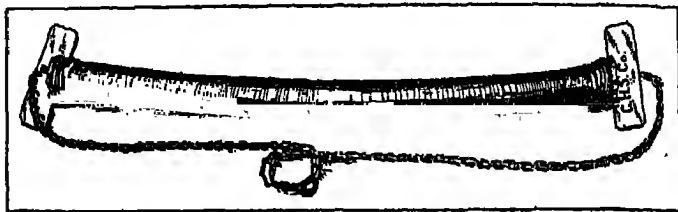
Breathing is another of the organic functions which shows marked change under conditions of excitement. Since breathing is the portal for the purification of the blood and its renewal with oxygen, it has an important place in the metabolic cycle. When a sudden and strong stimulus arouses the sympathetic nervous system to action, thus putting the body into condition for immediate and vigorous activity, among the other preliminary adjustments is a quickening of the breathing and an increase in its amplitude.

The rate of breathing may be noted simply by counting the number of inspiration movements of the chest during a minute or part thereof. This simple method, however, does not leave behind its own record. When a record of the rate of breathing is transferred to the kymograph, it is convenient at the same time to make a record of breathing amplitude, so the measurement of the two phenomena will be considered together.

The instrument for recording the characteristics of the respiratory movements is known as a *pneumograph*. The form usually employed was originally designed by Marey. In its simplest (and perhaps most satisfactory) form it consists of a stout, flexible rubber tube which may be stretched at will. This is

418 Diagnosing Personality and Conduct

stretched part way around the subject's chest, being held in place by a light chain and hook. One end of the tube is closed; the other end has a nipple, to which is attached the tube connecting the pneumograph with the kymograph tambour. As the subject breathes, the rubber tube is stretched, air being forced out. This change in air volume is transferred to the rubber membrane of the tambour.



PNEUMOGRAPH

A more complicated form of Marey's pneumograph consists of a plate of rather flexible metal so strapped to the subject's chest that it bends as the breath is inhaled and exhaled, thus causing changes in a Marey's tambour fixed directly above the plate by a series of levers. These changes in pressure are then transmitted by a rubber tube to a second Marey's tambour at the kymograph, where they are recorded.

The factors causing changes in respiration rate and amplitude are many. Although many of them are similar to those which cause changes in pulse, blood-pressure, blood volume, etc., others are different, so that a review of these factors will be helpful. There are no authoritative descriptions of conditions where rate and amplitude of respiration do not go together. Usually changes in one produce changes in the other.

The average rate of respiration is 18 per minute.

1. *Rate of respiration changes with age.* Table 81, taken from Burton-Opitz, gives the averages for different ages.

2. *Rate of respiration varies with the size of the animal.* In general the smaller the animal, the faster its rate of breathing. This is due to the fact that "smaller animals possess a more extensive body-surface in relation to their mass than the larger ones, and hence suffer a much greater loss of body heat."

3. *Changes in respiration are directly connected with changes in heart-beat, blood-pressure and the like.* Usually both are re-

TABLE 81

RATE OF BREATHING AT VARIOUS AGES

<i>Age</i>	<i>Rate of breathing</i>
New born	62
0-1	44
5-15	26
15-20	20
20-25	18.7
25-30	15
30-50	17

sults of the same cause, but there may be a direct reciprocal relation.

4. *Respiration is directly proportional to the CO₂ tension of arterial blood.* This is the primary factor producing momentary variations in breathing. Many of the factors which follow can be traced back to this influence of the blood composition.

5. *Rate and amplitude of respiration are increased with an increase in carbon dioxide content of the air.* The increase in the carbon dioxide content of the air must be considerable before any effect is noted. This is probably due to the inadequate washing-out of the carbon dioxide in the blood stream.

6. *Respiration increases with increased barometric pressure.* A corresponding increase in the depth of respiration is noted when the pressure is reduced in high altitudes. High barometric pressure as in caisson work, or low pressure as in airplane flights, produces a variety of physiological effects, of which change in breathing is only one.

7. *Rate and amplitude of breathing increase with exercise.* This phenomenon is too common for comment.

8. *Breathing is greater when erect than when reclining.* This is probably connected with muscular activity.

9. *Breathing decreases in sleep.*

10. *Breathing increases when speaking,* for obvious reasons.

11. *Breathing increases with increases in heat* both of the body and of the surrounding air. This is probably associated with the increased rate of metabolism in the body.

12. *Breathing increases momentarily when one is dashed with or immersed in cold water.*

13. *Breathing increases with pain.*

14. *Breathing may be increased or decreased in rate and amplitude by an act of will.*

420 Diagnosing Personality and Conduct

15. *Breathing is increased under the influence of emotion.*

Eng's (4) experimental work also yields data concerning the changes in breathing produced by various kinds of affective stimuli.

	<i>Respiration</i>
Attention-tension	Retarded
Psychical work	Accelerated
Displeasing taste and smell	Retarded
Pain	Accelerated
Fright	Retarded, followed by acceleration
Spontaneous displeasure due to sensory impressions	Small effect
Psychical displeasure (unpleasant concepts)	Small effect
Displeasure and excitement	Unchanged
Depression	Retarded
Pleasing smell	Accelerated and shallow
Spontaneous pleasure from sensory impressions or concepts	Accelerated

The same cautions that have been mentioned concerning pulse and blood-pressure apply to the use of changes in breathing as a measure of emotion. Only marked changes are significant, and it is not possible to make use of norms or the changes from one day to another, since all changes must be noted during the same sitting. Even so, extreme care must be used in not introducing any exciting change except that planned in the experimental procedure.

Measurement of the Psychogalvanic Reflex

Landis and De Wick (56), in an extensive review and summary of the literature bearing on the psychogalvanic reflex, tell us that the first mention of the "electrical tension of the skin" is to be credited to Bertholon in 1786, while Féré is ordinarily given the credit as being the first to demonstrate the psychogalvanic reflex in 1888. The early workers—Vigouroux, Féré, Tarchanoff, Sommer, and Veraguth—discovered most of the phenomena connected with the reflex and prepared theories in explanation of it. Vigouroux (98) in 1888 concluded that the electrical resistance of the body is a function of the vasomotor system. Féré (30) in the same year first called attention to the variations caused by

emotional changes. Tarchanoff (89) in 1890 attributed this electrical change to the secretory activity of the sweat-glands.

It has been found that if two points on the skin are connected so as to form an electric circuit, a current passes through them, due to a difference in potential at the two points on the surface of the body. This difference in potential is relatively large if one of the spots is a region of the skin rich in sweat-glands and the other a region relatively devoid of them; the difference in potential is small or absent for symmetrical points of the body, such as the corresponding spot on the two arms. Changes in mental activity, excitement, etc., will cause a change in the amount of current flowing through the circuit. If a small current be passed through the circuit from an outside source of constant voltage, the amount of this current will change with changes in emotional excitement. The electrical variations of the skin under emotional excitement without an exterior electromotive force is known as Tarchanoff's phenomenon. The same variations when an outside current is added is known as Féré's phenomenon. In the latter case the electrical changes due to emotional excitement are much greater. It is generally believed, however, that the changes in the two cases are both due to the same phenomenon.

In experimental work the amount of outside current to be used has been the subject of debate. Some have used none, although without an outside source of current the emotional effects are small and a correspondingly more sensitive galvanometer must be used. On the other hand, there is difficulty in maintaining a constant voltage and resistance in the system when an outside current is employed. The majority of experimenters use from two to three and one-half volts.

The amount of electrical current is measured by a *galvanometer*. There are many kinds of *galvanometers*. The kind generally employed in the extremely delicate work of measuring the electrical currents of the body is known as a *D'Arsonval galvanometer* (variously called a *moving coil galvanometer* or a *mirror galvanometer*). The principle on which this galvanometer is based is the relation between the current passing through a coil of wire suspended between the poles of a stationary magnet and the lines of force set up tending to move the coil. The coil of wire becomes an electromagnet when a current is passed through with poles at right angles to the poles of the stationary magnet. This sets

up a force tending to turn the coil and make the electromagnetic field in the coil become parallel to the field of the permanent magnet. In the D'Arsonval galvanometer there is a stationary horse-shoe magnet. Between the poles is suspended a coil of fine wire. The current enters the coil from below on a loose spiral wire, and leaves at the top by the fine wire on which the coil is hung. The coil is thus free to move between the poles of the magnet. Usually there is a stationary bar of soft iron in the space within



D'ARSONVAL GALVONOMETER

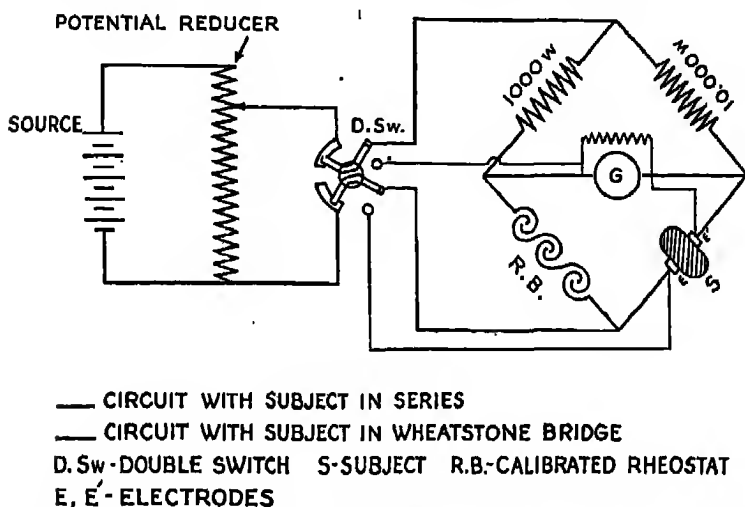
the frame on which the coil is wound to help concentrate the magnetic lines of force. The amount of twist or movement of the coil is a measure of the amount of current passing.

On the wire by which the coil is suspended is attached a mirror which is made to reflect the readings from a scale. If this mirror is viewed through a telescope having a fine vertical hair line, readings of the scale may be used as measures of the current. A beam of light may also be reflected onto a moving sensitive photographic film, thus giving a continuous record.

A Wheatstone bridge (an instrument for measuring electrical resistance) is usually placed in the circuit, both in order to meas-

ure the resistance offered by the body to the passage of the current, and also for the purpose of bringing the galvanometer needle to zero at the beginning of the experiment. The Wheatstone bridge is not necessary, however, to the experiment.

Wechsler (110) found that the electrical resistance of the body varies during the progress of an experiment, often in an unpredictable way. He believed it to be desirable to make some sort of adjustment of the *exsomatic* (from outside the body) current.



ELECTRIC CONNECTIONS FOR APPARATUS FOR MEASURING THE PSYCHOGALVANIC REFLEX

From Wechsler, David, *The Measurement of Emotional Reactions*, Archives of Psychology, No. 76, p. 127.

This he accomplished by a *potentiometer* (a rhcostat which permits variations in the amount of current allowed in the circuit and inadvertently, as it were, measures the amount of current), which is manipulated during the process of the experiment when it is noticed that the galvanometer needle does not go back to zero when at a state of rest. This does not seem to the writer to be defensible experimental technique. If such a procedure is necessary, it is testimony to the grave difficulty of using the psychogalvanic reaction as a measure of emotion at all.

Godefroy (41) employed an interesting circuit which includes a *transformer*. The subject is in circuit in the primary coil, while

the galvanometer is in the secondary coil circuit. In this circuit the galvanometer needle goes back to zero of itself after every change of the current, so that variations in the body resistance do not act as an influence on the galvanometer. This is a happy solution of the vexing problem of allowing for variable resistance.

The latest device for measuring small electrical currents very sensitively is the *cathode-ray oscillograph*, an instrument with which alternating current must be employed. The most simple of these instruments is the *Brown tube*. This is an evacuated glass bulb. A cathode ray is caused to pass through the tube, striking at the further end of the tube a zinc sulphide layer which becomes illuminated when the rays strike it. Two plates are placed on opposite sides of the tube and connected in the psychogalvanic circuit. As the current passes through these plates, they *bend* the electric stream and cause the spot of light on the zinc sulphide disk to move. This spot of light can be picked up photographically, whence it becomes a measure of the current passing through the body.

The question of the *electrode* to be used in the psychogalvanic reflex experiment is very difficult. (An *electrode* is the plate or terminal of an electric current. In this case it is the material used to make the contact with the body which completes the electrical circuit.) Wechsler gives three conditions to be considered in the choice of electrodes: (a) the part of the skin to which they are applied, (b) the security of the contact, (c) the degree of polarization of the electrode. All electrodes *polarize*. (By *polarization* is meant the accumulation on the surface of the positive electrode of some substance (a layer of ions) which increases the resistance in the circuit. This polarization commences the moment the circuit is completed and the current commences to pass, and acts continuously to increase the resistance.) A variety of electrodes have been tried in the experimental work. Some workers have used dry metal plates of brass, zinc, copper, tin, bronze, etc. Others have tried liquid solutions. Wechsler had his subjects immerse two fingers of the same hand in porous clay cups filled with salt solution and containing small plates of amalgamated zinc. The clay cups in turn were set in small glass jars filled with zinc sulphate. The size of the electrodes also has been given experimental consideration. In general it has been found that the greater the electrode surface, the less effective is the polarization.

Several interesting features of the psychogalvanic reflex may be noted. For the first ten or fifteen minutes after the electrical circuit is closed, there is a diminution or fall in resistance, so that in experimental procedure it is well to wait until the galvanometer needle has come to rest. This phenomenon is said to be due to an initial polarization of the skin. A second feature of the reflex is a latent period of from two to three seconds after a stimulus is given. A third characteristic of the psychogalvanic response is a small negative deflection which usually precedes the main deflection. The presence of this phenomenon when all conditions for the reflex are right is doubted by some investigators, who attribute it to poor electrode contact, slight muscular contractions in the region of the electrode, or vaso-dilation. Another unexplained phenomenon is a small negative deflection following a positive deflection, a phenomenon which is so small, however, as never to be serious in practical work.

Explanation of Psychogalvanic Phenomena

That the skin should possess electrical properties is most interesting, and naturally many theories have been advanced to explain the phenomenon. After their careful and exhaustive review Landis and De Wick (56) conclude that, "Conflicting evidence forces us to wait for further, better controlled, and more scientific experimentation before accepting any definite physiological explanation." Three theories, in general, have been advanced in explanation of the psychogalvanic phenomena, which may be called respectively the *sweat*, the *vasomotor*, and the *muscular* theories.

Jeffress (48, p. 141) is the best exponent of the *sweat* theory, which he describes as follows:

"If we consider the interior membrane of the sweat-glands as semipermeable, we may describe the phenomena of secretion as an alteration of permeability which permits the passage of the perspiration from the interior of the sweat-glands into the tubules. If we further assume that the membrane is permeable to only one of the ions in the sweat, which is already in the tubules, or in the electrolyte of the electrodes, we may explain the phenomena of polarization through the skin. If the membrane is permeable to only one of the ions, say the cation, the others collect at the membrane and oppose the flow of current; in other words, the membrane becomes polarized and a counter-electro-

motive force is set up opposing the flow of current. Now if upon stimulation (nervous), the permeability of the membrane is changed so that it is more or less permeable to both ions, the current would flow with less difficulty; i.e., the counter-electromotive force would be reduced. This, as we have seen, is what actually occurs upon stimulation."

Jeffress goes on to show that the nervous innervation of the sweat-glands is itself an electrical phenomenon (48, p. 142): "We should expect, then, that stimulation of the sweat glands would exhibit itself by an electrical manifestation or 'action current' of the gland itself. Several experiments (Waller and Mendelssohn) have shown that activity of the sweat-glands is accompanied by an increase in electrical potential directed from the surface of the skin inward."

Jeffress himself found that the hand, which is relatively rich in sweat-glands, becomes more strongly negative to the indifferent parts of the body (mouth) when the subject is stimulated, leading him to believe that it is a case of the action currents of the sweat-glands.

This theory has much to commend it.

The *vasomotor* theory has its strongest champions to-day in McDowall and H. M. Wells (19, 20, 59, 113, 114). These experimenters have demonstrated a very close correspondence between the constriction of the blood vessels of the skin and the psychogalvanic reflex. On the other hand they cite cases from their data in which profuse sweating under the influence of drugs yielded only a normal needle deflection. The evidence is sufficient to indicate that there is a very close connection between the vasomotor system and the psychogalvanic reflex, but this connection may be correlative only, since changes in the vasomotor system accompany excitement changes in general. However good the grounds for believing that the vasomotor system may be in part responsible for the psychogalvanic reflex, we cannot at present understand the connection.

The *muscular theory* seems less tenable, and has not been stressed in the most recent work.

Factors Influencing the Psychogalvanic Reflex

The use of the psychogalvanic reflex for the measurement of emotion is beset by difficulties, for besides the outside stimuli

which cause physiological variations, the factors which cause variations in an electrical current must be considered.

1. *Variation in the bodily resistance* is one of the major difficulties in using the psychogalvanic reflex for the study of emotion. The electrical resistance of the body varies in an unpredictable way during the course of an experiment, and quite apart from changes caused by the experimental stimuli presented.

2. *The polarization of the electrodes* influences the needle deflection. The greater the polarization, the greater the resistance and the less the deflection. Relatively unpolarizable electrodes are the most satisfactory.

3. *The size of the electrode* influences the needle deflection by making polarization less effective.

4. *The points on the skin where the electrodes are applied* have a marked effect on the needle deflection. In general the reflex is more pronounced when one of the electrodes at least is connected with a region rich in sweat-glands.

5. *The security of contact of the electrodes influences the size of the deflection.*

6. *The intensity of the current influences the size of the reflex.* Within limits, the stronger the exsomatic current used, the greater the extent of the galvanometer needle deflection. It is essential, therefore, that the strength of the current be constant during an experiment, unless proper allowance and correction are made.

7. *Muscular movement influences the reflex.* Many investigators have noted the relation between muscular activity and the psychogalvanic reflex. Indeed some have tried to see a direct causal connection between muscular activity and the electrical resistance. One can understand, however, how muscular activity has its influence on the peripheral blood vessels and the sweat-glands through the sympathetic nervous discharge, and it seems unnecessary to hypothesize a direct connection between muscular activity and the reflex.

8. *The length of time the exsomatic current flows has an influence on the reflex.* Wechsler (110) gives evidence to show that in cases when a low intensity current was used with an electrode placed on a spot rich in sweat-glands (the palm of the hand), the resistance increases with time. This is probably a polarization effect. On the other hand, when high intensities of current were tried, the resistance fell off with time, which Wechsler attributes

to inhibition (absorption). Between these two extremes there is every possible variation. Wechsler finds that with low-voltage currents applied to the finger-tips, the two opposing tendencies are neutralized and the resistance remains more nearly constant.

9. Peterson (62), Binswanger (24), and Moravcsik (60) note an *increased deflection when pressure is applied to the electrodes*. H. M. Wells (59) on the other hand reports that pressure on the electrodes may diminish or even prevent the reflex. Further research is needed on this point.

10. *Resistance to alternating current is less than to direct current*, and does not produce such large fluctuations.

11. Gildemeister reports that the *positive variation* of a constant current passed through the skin on stimulation of the skin surface is *accentuated by raising the temperature of the electrodes*, and that this local warming increases both the local time and the duration of the reflex. Landis (55), on the other hand, reports *no relationship between temperature changes of the skin-electrode picture and the resistance of the body*. It is conceivable that, if the sweat-glands are stimulated to action by heat, the reflex could increase with a rise in general surrounding temperature. Sidis and Kalmus (77) find that heat and cold have little effect on the psychogalvanic reflex, while Wells (114) reports that it is difficult to elicit the reflex in cold weather. Evidently further investigation must be looked to before this point can be settled.

12. *Fatigue diminishes the reflex*. In this connection Wechsler's summarized statement of the diurnal variations of the reflex may be cited: "The human body resistance tends to be least toward the middle hours of the day and greatest at night and in the early morning hours." The lessened resistance during the day may be attributed in part to muscular activity and the increased resistance in the evening perhaps may be explained as a fatigue phenomenon. No explanation can be given for the high resistance in the early morning.

13. Closely related to the foregoing is the *decrease in the galvanometric deflection caused by repeated stimulation of sense-organs*.

14. *Deep breathing provokes a psychogalvanic reflex*. This factor probably operates through the sympathetic nervous system, leading to changes in the surface arteries and sweat-glands.

15. *Mental activity causes a diminution of the psychogalvanic reflex*, thus acting in the direction opposite to that of muscular activity. On the other hand hard mental work accompanied by strain or effort increases the reflex.

16. *Any direct increase in stimulation of the sweat-glands stimulates the psychogalvanic reflex and vice versa.* Certain instances, exceptions to this rule, are noted in the experimental literature. Wells (59) reports a case in which profuse sweating caused by drug action was actually accompanied by a decreased psychogalvanic reaction. Such cases are rarely reported, however, and the testimony is overwhelming that the reflex is positively correlated with sweat secretion.

17. *Drugs have a marked influence on the psychogalvanic reflex*, but here the reports in the experimental literature are extremely conflicting.

18. Many investigators have reported that the *psychogalvanic reflex can be elicited by sensory stimulation alone*. Starch (82) and Veraguth (92) believe that the amplitude of the needle deflection corresponds roughly to the strength of the stimulus. Since repeated stimuli cause the reflex to become less, it is evident that the psychogalvanic reflex follows the usual laws of attention and fatigue.

19. *The evidence as to whether the psychogalvanic reflex can be voluntarily controlled is conflicting.*

20. Finally *emotional changes which cannot be shown to have somatic correlates cause a psychogalvanic deflection*. It would seem as though, by the time all of the foregoing factors had their influence on the psychogalvanic reflex, that the deflections due to the presence of emotion would be quite lost and undistinguishable. The psychogalvanic reflex is, of course, commonly spoken of as a measure of emotional reactions. The point is that during a brief period in the laboratory when all external disturbances are eliminated and all physiological factors are kept constant so far as possible, galvanometric deflections appear which are probably closely correlated with emotional changes. The warnings must be repeated that we should be wary in comparing experimental results when the experimental situation varies in any significant way, and that it is safest never to compare the results of two separate sittings.

Practical Use of Physiological Measures of Emotions

Physiological measurement of the emotions could be employed in a wide range of situations for the purpose of determining the extent to which various situations arouse emotional responses. Such measures might be used in the school-room to discover the degree to which teachers, contests, reciting before the class, examinations, and different methods of instruction or of control arouse emotional response. They might be used in the home to discover the influence of various situations and events. They could be used in the theater, the motion-picture house, in industry, in athletic contests—in fact every phase of life awaits testing of the arousal of emotional excitement. Although one cannot generalize and although one certainly would not recommend that all emotional reactions be eliminated, it is true that a response which is accompanied by emotion is generally not the most satisfactory form of adjustment. We look hopefully to the future therefore for exact knowledge of the situations which normally arouse emotions.

The Psychogalvanic Reflex and the Association Experiment

Landis and De Wick (56) are authority for the statement that Veraguth and Clotta were the first to point out the relationship between the psychogalvanic reflex and reaction time. Jung (50), however, in 1907 was the first to capitalize the discovery by finding that the psychogalvanic reflex in connection with the response in the free association experiment was an index of an underlying complex or area of emotional response. F. Peterson (62) believes that the galvanometer deflection is directly proportional to the degree of emotion aroused. W. W. Smith has studied this relationship extensively. As a result of his work he believes the psychogalvanic reaction is a better index of the affective nature of the response than is reaction time. He says (80, p. 74):

"There can be no doubt whatever that for quantitative work the galvanometer deflection is a far more valuable indication than the reaction time. It is not under voluntary control and is not affected to any appreciable extent by non-significant intellectual factors such as sometimes prolong reaction time. Moreover, the absolute magnitude of the deflections can, in general, be magnified

to any extent desired and read with a corresponding degree of precision. Still more important is the fact that the magnitude of the galvanometer deflection appears to be approximately proportional to the intensity of the corresponding affective tone, however great the deflection."

As evidence for the latter statement Smith presents two tables giving the rank order of 100 free association words according to reaction time and according to galvanometer deflection. The six words resulting in the highest reaction time are *name*, *friend*, *despise*, *make*, *sad*, *proud*, and *home*; while the six words giving the greatest galvanometric deflection are *kiss*, *love*, *marry*, *divorce*, *name*, and *woman*. It is not difficult to estimate that the affective quality of the second group of words is greater.

The psychogalvanic reflex promises to be of service in interpreting the reaction experiment by indicating whether failure to respond or a long reaction time is due in the particular case more precisely to intellectual difficulty in interpreting the word or to a conflict of two or more innocuous stimuli, or is really accompanied by emotion. It might also be serviceable in the case of a prolonged reaction time to indicate whether or not implicit reactions with an emotional coloring are aroused, though they pass without overt expression.

The greater number of investigators trying the psychogalvanic reflex on patients with mental disorders have concluded that the reflex cannot be used with certainty in diagnosing mental disorder. The testimony of various investigators is very contradictory. Some claim that the reflex can help in differentiating between hysteria and hysterical simulation, or between functional and organic mental disorders, while others insist that the reflex is without power to distinguish between these cases. Syz (87) found that patients in catatonic stupor have an average electrical resistance twice as high as normal persons, but such findings are not helpful in diagnosing the nature of the difficulty, as other mental states also result in high electrical resistance.

Use of Physiological Measures of Emotion in the Detection of Deception

Most of the work in which practical use has been made of the various physiological measures of emotion has been done in the

attempt to determine whether or not a person is lying. To the psychologist this seems a rather picayunish problem. From the point of view of science it appears of comparatively small consequence that in a given situation a person has told the truth or a falsehood. Of more consequence is the knowledge whether a person has some emotional complex or some compulsive tendency or interest in committing a crime or other unsocial act.

But in practical affairs there is often a strong desire to determine the facts in a situation where deceit may be involved. In legal criminal procedure at present, where the whole emphasis is on conviction for a *particular* offense, there is a pressing practical need for some device which will give a check on the veracity of testimony and confessions. So those who have worked in this field—Benussi (117) and Lombroso (127) some years ago in Italy, and more recently Marston, Larson, and Landis in this country—have approached the problem with an interest in its applications to police or legal procedure. The contemporary workers have held with naïve enthusiasm to a belief that methods will soon be devised, which will guarantee beyond doubt that a person is lying or telling the truth. Marston (129), using systolic blood-pressure, claims an accuracy ranging from 90 to 100 per cent. Larson (122, 123) claims over 90 per cent accuracy in his work with a combination of blood-pressure and breathing records. (How much over 90 per cent cannot be known because of disappearance of a subject, "inability to obtain a confession" (!), and other causes.)

The reported experimental work, with the exception of that of Landis, is extremely unsatisfactory. Altogether too much has been left depending on the individual judgment of the experimenter. Most of the work has been done from the point of view of being able to spot the individual liar rather than of systematically investigating the relationships between the various measures used and telling the truth. The statistical handling of the data collected has been crude beyond words. The amount of deviation in any measure which distinguishes truth from falsehood has been established arbitrarily. In short, the experimental work to date is mainly rough pioneering.

The theory underlying the work has been that when one tells the truth there is no particular arousal of the emotions, but that when one is deliberately concealing the truth by falsifying, there

is an accompanying emotional tension or strain. When telling the truth the tendency is toward relaxation; when lying, one is on guard, and there is a definite heightening of effort and tension. Of course the hardened criminal who is so obtuse or so practised in his art that he can lie without "batting an eye" would show no such emotional disturbance, but such an individual is comparatively rare, according to the belief of those who have worked in the field. Nearly every one can draw on his own experience for testimony concerning the accompaniments of a lie—the racing heart, the gasp in the breath, increased perspiration, a dry mouth, and the like. If one could be certain that these things always happened, and if they could be detected experimentally, then we might attain a sure-fire diagnosis of deception.

Most of the experimental work has been carried on in the laboratory or college class-room with artificial "crimes" committed by members of the class. Some work has been done in police courts with actual criminal suspects. It is the belief of those who have worked in this field that the laboratory situation is not really a fair trial for the method. When the situation is not real, the emotional effects may be inhibited, but when a criminal suspect is really attempting to defend himself before society, the emotion is hard to keep down.

Although any of the physiological measures of the emotions previously described might be used for this purpose, the two that have been most commonly tried are (a) systolic blood-pressure and (b) breathing.

Marston (129) waxes enthusiastic over the possibilities of blood-pressure. Trying it out first in faked-up crimes in the laboratory and later in actual criminal cases during and after the World War, he achieved remarkable success. In laboratory experiments Marston reports 96 per cent correct judgments of truth or falsehood on the basis of blood-pressure records, whereas a "jury" who listened to the questions and answers ranged from 10 to 85 per cent correct with an average of 48 per cent. Again, during the World War, Marston (128) reported 94.2 per cent correct judgments by himself, while privates in the United States Army, given the same blood-pressure records to interpret, were able to make correct judgments in only 74.3 per cent of the cases.

Marston claims that there are typical blood-pressure curves of truth and falsehood. The curve of falsehood shows a steady

434 Diagnosing Personality and Conduct

rise as the various questions are asked, reaching a peak at the crisis in the testimony. The curve of truth, on the other hand, pursues a level course with few deviations. Naturally one must distinguish between the changes in the curve due to the presence of normal complexes and those due to lying. Marston believes that lying always produces a greater rise in the blood-pressure curve than any incidental emotional complex. In his early work he set eight millimeters' rise in the mercury level of the sphygmomanometer as symptomatic of lying, but later he raised this figure to twelve millimeters.

Larson (123, p. 450) notes the following as indices of deception:

1. Increase in blood-pressure.
2. Decrease in blood pressure.
3. Increase in height (pulse?).
4. Increase in frequency.
5. Summative effects.
6. Incomplete inhibition (breathing?).
7. Complete inhibitory effect.
8. Irregular fluctuations, especially noticeable at the base of each cardiac pulsation.
9. Combination of any of the above effects in the same individual.
10. These changes may occur with but little latent period, or they may be accumulative in effect and more generally distributed.

Landis (119, 120), in collaboration with Gullette and Wiley, gave the systolic blood-pressure method of detecting deceit a careful trial and found that he was able to diagnose correctly only twelve out of twenty-two cases. That is only slightly better than chance. Landis (121) concludes, "For this group and with this technique, blood-pressure reactions were an untrustworthy criterion of consciousness of deception." But in another place Landis admits that his experimental conditions may account for his poor results and says, "We are of the opinion that the blood-pressure method of detection of falsehood is what Marston originally claimed for it, 'highly diagnostic,' if all conditions are favorable."

In using breathing for the detection of deception, the recent experimenters have gone back to a discovery made by Benussi that under the influence of a strong emotion the time for inspira-

Physiological Measures of the Emotions 435

tion is greater than the time for expiration. When one laughs there is a tendency to take the breath in quickly and then let it out more slowly in little short gasps. In anger or fear, on the other hand, one breathes out rapidly, the inspiration is more regular, and the breath is held after it is taken in. Breathing is measured and recorded by means of a pneumograph, as described earlier in this chapter. Feleky (5) has determined the following inspiration-expiration ratios (I/E).

TABLE 82
INSPIRATION-EXPIRATION RATIOS FOR VARIOUS STIMULI
(from Feleky)

Normal breathing805
Laughter30
Hatred515
Disgust	1.08
Pleasure	1.11
Pain	1.546
Anger	1.48
Wonder	2.49
Fear	2.66

Benussi (117) claims practically 100 per cent certainty in using the inspiration-expiration ratio in the determination of deception. Although breathing generally is under voluntary control, the I/E ratio under the influence of emotion seems to be practically independent of intentional control. Certain improvements in the method of recording the I/E ratio directly have been described by Burt (12, 13). He was successful (118) in detecting the truth or falsehood in 73 per cent of the cases which he tested. He believes, however, that blood-pressure is a more accurate index than the I/E breathing ratio, partly because he notes that the influence of lying upon breathing appears to become less with habituation. Larson claims exceptional success in using the inspiration-expiration ratio in conjunction with records of pulse and blood-pressure.

Landis (120), who with Wiley approached the problem experimentally and without evident bias, found that the I/E method gave diagnostic results in 63 per cent of the cases when subjects were questioned concerning the figures on a card and in 50 per cent of the cases when they were questioned concerning an imaginary crime. Landis states, "We may conclude that we are dealing with a real factor which does vary from truth to falsehood."

436 Diagnosing Personality and Conduct

It is evident that one cannot correctly diagnose deception unless all the conditions are favorable. Some of these conditions are as follows:

1. *Perfection of instruments.* Defective or carelessly set up instruments will cause any laboratory experiment to fail, as will careless manipulation of the machinery. Laboratory apparatus of the sort we are describing is not at present adapted to rough usage.

2. *Reality of the situation.* There is evidence for believing that the ordinary demonstration before a psychology class of the experiment with planted and artificial crime does not produce the emotional shock necessary to give the method a fair trial. Actual cases in the police court or "third degree chamber," where the defendant is driven to the necessity of proving his innocence to avoid conviction, possess a reality which should cause emotional upheavals if anything can.

3. *Familiarity with the procedure.* There is some evidence that the results are better when the subject is unfamiliar with the procedure. If the subject knows that the complicated apparatus strapped on his arm and around his chest is measuring his heart action and breathing he may make a partially successful effort to control his emotions. On the other hand, if the subject is totally unaware of the significance of the experiment, his replies to the questions asked will be accompanied by all the usual emotional disturbance.

4. *Types of questions.* Little is known of the emotional influence of different types of questions. This is a fertile field for investigation and one of the highest importance for criminal psychology. It is not known, for instance, whether the emotional effect is greater in a room with an atmosphere of excitement and tenseness in which there are several people present than in a quiet room alone with the experimenter. It is not known whether it is more effective to plant significant questions casually between neutral questions, or to proceed with a string of direct leading questions bearing on the crime, or to arrange the questioning so as to lead up to a dramatic climax. It is not known whether the questions should be put in a short space of time or whether the examination should be strung out over a protracted interval. A recent newspaper article describing the use of the "lie-detector" in forcing a confession made mention of continuing the inquisition over several days. If it was continued too long, one might expect anger

to be aroused in an innocent person by mere impatience at being confined or insulted unnecessarily.

5. *Emotional susceptibility of the subject.* It is well known that some persons are far more sensitive to telling a lie than others and that there are large individual differences in the emotional effect of telling a falsehood.

There is immense popular interest in methods of detecting deception. A great deal of further careful experimental work must be done, however, before the method can be considered ready for practical use.

Biochemical Measurements and Personality

Rich (136, 138) has discovered certain rather surprising relationships between the acidity of saliva and urine on the one hand and personality characteristics on the other hand that offer a promising lead in the diagnosis of personality. By making a determination of the hydrogen-ion concentration of the saliva, he found a positive relation between the *alkalinity* of the saliva and ratings for *excitability*. That is, since hydrogen-ion concentration decreases as a liquid becomes less acid, the most excitable individuals have the least acid saliva and vice versa. Starr (139) earlier found the same relationships when working with two types of stammerers.

Corresponding facts were found in testing the acidity of the urine, which correlates negatively with emotional excitability. The most excitable individuals have the least acid urine and vice versa. Similar relationships were found in correlating with rating for aggressiveness. The following table gives the findings.

TABLE 83
CORRELATIONS OF ALKALINITY OF SALIVA AND ACIDITY OF URINE WITH
PERSONALITY RATINGS

(from Rich)

	39 UNDERGRADUATES		18 GRADUATES	
	<i>Aggres- siveness</i>	<i>Excit- ability</i>	<i>Aggres- siveness</i>	<i>ability</i>
Salivary alkalinity	+ .09	+ .28	+ .51	+ .45
Urine acidity	- .31	- .25	.00	- .26
Urine acidity (formal titration).	- .31	- .06	- .24	- .27
Urine acidity (formal titration) divided by body weight.....	- .49	- .22	- .31	- .17

438 Diagnosing Personality and Conduct

The correlations indicate that bodily alkalinity goes with aggressiveness and excitability, whereas bodily acidity goes with passivity and lassitude.

Determination of Amino-Acid Nitrogen

The waste products of metabolism contain a number of amino-acids. In ordinary urine these are approximately neutral in reaction. If, however, formaldehyde is added, a reaction takes place in which the basic properties of the amino groups are destroyed, leaving carboxyl groups which are more strongly acid. These may be titrated against an alkaline solution and the acidity determined. The method commonly used is known as the Henriques-Sorenson or Formal Titration Method. Rich added the acidity as determined by direct titration to the acidity as determined by the formal titration to give a measure of total acidity.

The amount of acid found in a twenty-four hour specimen of urine has a direct relation to the weight of the individual, since the heavier an individual the more tissue there is which is being metabolized, and the larger the amounts of waste substances excreted in the urine. Consequently Rich divided the urine and acidity findings in each case by body weight to eliminate what might prove to be a disturbing factor in his comparisons.

The creatinine concentration of the urine was also found by Rich (131) to be associated with personality factors. Since creatinine is a product of the process of normal metabolism taking place in the muscles which is finally eliminated through the urine, the creatinine output in the urine is a rough measure of the metabolic process.

Rich found that the creatinine concentration of the urine correlates $-.24$ and $-.10$ with ratings of emotional excitability for eighteen graduate and thirty-nine undergraduate students respectively. When the creatinine output is divided by the body weight (creatinine coefficient), the corresponding correlations with emotional excitability are $-.23$ and $-.24$. This small but consistent relationship indicates that here is a factor of some importance in determining personality trends. Less excitable individuals tend to have a more rapid creatinine metabolism and vice versa.

The method usually employed for determining the creatinine in urine is that developed by Folin and is called Folin's Colori-

Physiological Measures of the Emotions 439

metric Method.* This method is based on the characteristic property of creatinine, that of yielding a certain definite color-reaction in the presence of picric acid in alkaline solution.

Each of these measures of metabolism varies according to certain factors.

The acid content of the saliva varies according to the *bacterial activity* between the crevices of the teeth, which tends to make saliva acid, and according to other obscure factors related to *metabolism*.

The acidity of the urine varies according to the influence of a number of factors, among them being:

1. *Diet.* Since diet may be one of the factors which determine excitability or aggressiveness, as well as the character of the urine, Rich made no effort to control diet, but let his subjects eat as they usually did. His only restriction was to omit certain articles of food, such as liver, that have a rather marked influence on the composition of the urine and would cause variations outside of the normal range.

2. *Water intake has a decided influence on the concentration and constitution of the urine.*

3. *Muscular activity influences the acidity of the urine.*

4. *The acid products in the urine are a function of bodily weight.* The heavier a man, the more muscular tissue there is to be metabolized.

5. *Acidity of urine tends to be increased by certain pathological disorders.*

6. The acidity of the urine is a *function of the time of day when it is passed*. For some time after a meal the urine may even give an alkaline reaction. This phenomenon is called the *alkaline tide* and is due to the chemical process of digestion.

7. *Urine may become alkaline, or even more acid, upon standing for some time*, due either to bacterial decomposition or to fermentation.

The creatinine concentration of the urine is said not to be affected by the amount of muscular activity over twenty-four-hour periods.

1. Creatinine output is dependent in part on the nature of the *food taken*.

* A description of the procedure may be found on page 732 of Hawk and Bergeim, *Practical Physiological Chemistry*, ninth edition.

2. *Creatinine output is a direct function of the metabolic process of the muscles* in such a way that the muscular efficiency of an individual depends upon the intensity of the process.

3. *Creatinine output changes under pathological conditions.* It is said to increase in typhoid fever, typhus, tetanus, and pneumonia and decrease in anemia, chlorosis, paralysis, muscular atrophy, degeneration of the kidneys, and leukemia.

Rich also determined the alkalinity and creatinine content of the blood, but since these methods should be undertaken only by one specially trained in blood-letting, they will not be described here. The same relationships, however, apply to the character of the blood that were found to be true for the saliva and urine.

One can only speculate as to the possible causes of the relationship between these physiological factors and personality factors. It is known, for instance, that glandular secretion is rendered easy or difficult by a change in the acidity of the blood. An increase in the acid character of the blood tends to retard glandular secretion. If these personality differences are due to extent of the secretion of the adrenal or pituitary or other glands of internal secretion, and the activity of these glands is directly related to the acidity of the bodily tissues, then the explanation of Rich's findings is clear. But this is only a hypothesis.

Again the creatinine excretion may be a measure of glandular activity. If glandular products act as catalyzers in hastening or retarding metabolism, this fact would be shown by the products of metabolic activity. Consequently if emotional excitability is also a function of glandular activity, it may be measured in this round-about way by the products of metabolism. These again are only speculations.

Chemical determinations of personality are certain to loom importantly in future work, which may uncover important facts bearing on the causes of personality differences.

REFERENCES

Physiological Changes under Emotion

1. ANGELL, J. R., and THOMPSON, H. B., "Organic Processes and Consciousness," *Psychological Review*, 6: 32-69 (1899).
2. BLATZ, W. E., "The Cardiac, Respiratory and Electrical Phenomena Involved in the Emotion of Fear," *Journal of Experimental Psychology*, 8: 109-132 (1925).

Physiological Measures of the Emotions 441

3. COLEMAN, W. M., "On the Correlation of Rate of Heart Beat, Breathing, Bodily Movement and Sensory Stimuli," *Journal of Physiology*, 54: 213-217 (1920-1921).
4. ENG, H., *The Emotional Life of the Child Compared with that of the Adult* (Oxford University Press, 1925).
5. FELEKY, A. M., "The Influence of Emotions on Respiration," *Journal of Experimental Psychology*, 1: 218-241 (1916).
6. HYDE, I. H., and SCALAPINO, W., "The Influence of Music upon Electrocardiograms and Blood Pressure," *American Journal of Physiology*, 46: 35-38 (1918).
7. LANDIS, C., GULLETTE, R., and JACOBSEN, C., "Criteria of Emotionality," *Pedagogical Seminary and Journal of Genetic Psychology*, 32: 209-234 (1925).
8. LANDIS, C., "Studies of Emotional Reactions. V. Severe Emotional Upset," *Journal of Comparative Psychology*, 6: 221-242 (1926).
9. LUDLUM, S. D., "Physiologic Conditions Under Which Insanity Occurs," *Archives Neur. and Psychiat.*, 11: 282-291 (1924).
10. SKAGGS, E. P., "Changes in Pulse, Breathing and Steadiness under Conditions of Startledness and Excited Expectancy," *Journal of Comparative Psychology*, 6: 303-317 (1926).
11. TOTTEN, E., *Oxygen Consumption During Emotional Stimulation*, *Comparative Psychology Monographs*, 3: 13 (1925).

Techniques for Measuring Blood-Pressure and Inspiration-Expiration Ratios

12. BURTT, H. E., "Further Technique for Inspiration-Expiration Ratios," *Journal of Experimental Psychology*, 4: 106-110 (1921).
13. ———, "A Pneumograph for Inspiration-Expiration Ratios," *Psychological Bulletin*, 15: 325-328 (1918).
14. ERLANGER, J., "A New Instrument for Determining the Minimum and Maximum Blood Pressures in Man," *The John Hopkins Hospital Reports*, 12: 53-110 (1904).
15. KOLL, A. C., "Continuous Blood Pressure Tracings in Man," *Journal of Phar. and Exp. Therapeutics*, 15: 433-441 (1920).
16. LANDIS, C., "A Method of Obtaining Records of Constant Blood Pressure," *American Journal of Psychology*, 34: 470 (1923).
17. NIXON, H. K., "A Simple Device for Recording Inspiration-Expiration Time," *Journal of Experimental Psychology*, 6: 383-386 (1923).

The Psychogalvanic Reflex

18. AVELING, F., "The Conative Indications of the Psycho-Galvanic Phenomenon," *Proc. Eighth International Cong. of Psychology*, 8: 227-234 (1926).
19. AVELING, F., and McDOWALL, R. J. S., "The Effect of the Circulation on the Electrical Resistance of the Skin," *Journal of Physiology*, 60: 316-321 (1925).
20. AVELING, F., McDOWALL, R. J. S., and WELLS, H. M., "On the Physiology of the So-called Psycho-Galvanic Reflex," *Journal of Physiology*, 60: vii-viii (1925).
21. BAYLISS, W. M., "Changes in the Electrical State of the Body and Their Relation to the Nervous System," *Brain*, 40: 536-549 (1917).
22. BIJTEL, J., and VAN ITERSON, C. J. A., "Psychogalvanic Reflex Phenomena in Sense Organs, Especially the Nose," *Acta Oto-Laryngologica*, 7: 30-40 (1924-1925).
23. ——— and ———, "Recherches sur le réflexe psycho-galvanique, en particulier comme réaction aux réactions aux excitations olfactives," *Arch. néerl. de Physiol.*, 10: 447-448 (1925-1926).
24. BINSWANGER, L., *On the Psycho-galvanic Phenomena in Association Experiments* (Moffat, Yard & Company, 1919).
25. BROWN, W. S., "A Note on the Psycho-galvanic Reflex Considered in Conjunction with Estimates of Character Qualities," *British Journal of Psychology* (Gen. Sect.), 16: 130-135 (1925).
26. DARROW, C. W., "Sensory, Secretory, and Electrical Changes in Skin Following Bodily Excitation," *Journal of Experimental Psychology*, 10: 197-226 (1927).
27. DUNLAP, K., "Galvanometric Deflections with Electrodes Applied to the Animal Body," *Psychological Bulletin*, 7: 174-177 (1910).
28. ———, "The Galvanic Phenomena," *Psychological Bulletin*, 8: 140-144 (1911).
29. ———, "Psychogalvanic, Circulatory, and Respiratory Phenomena," *Psychological Bulletin*, 10: 142-146 (1913).
30. FÉRE, C., "Note sur des Modifications de la Tensions Électrique dans le Corps Human," *Compt. Rend. Soc. de Biol.*, 5: 28-33 (1888).
31. ———, "Note sur les Modifications de la Résistance Électrique sous l'Influence des Excitations Sensorielles et des Émotions," *Compt. Rend. Soc. de Biol.*, 5: 217-219 (1888).
32. FLEMMING, E. G., "Personality as Revealed by the Galvanometer," *American Journal of Psychology*, 38: 128-129 (1927).

Physiological Measures of the Emotions 443

33. FOA, C., and PESERICO, E., "Le vie del riflesso neurogalvanico," *Arch. di Fisiol. Firenze*, 21: 119-130 (1923).
34. GILDEMEISTER, M., "Über die physikalisch-chemischen und physiologischen Vorgänge im menschlichen Körper auf denen der psychogalvanischen Reflex beruht," *Münch. Med. Wochenschr.*, 60: 2389-2391 (1913).
35. ———, "Über eine Veränderung von Zellmembranen unter nervösen Einfluss," *Arch. Internat. de Physiol.*, 14: 2-19 (1913-1914).
36. ———, "Über electrischen Widerstand, Kapazität und Polarisation der Haut," *Pflüger's Arch.*, 176: 84-105 (1919).
37. GILDEMEISTER, M., and KAUFOLD, R., "Über das elektrische Leitungvermögen der überlebenden menschlichen Haut," *Pflüger's Arch. f. d. g. Physiol.*, 179: 154-158 (1920).
38. GILDEMEISTER, M., "Zur Theorie der sekundär-elektromotischen Eigenschaften der menschlichen Haut," *Pflüger's Arch.*, 195: 112-122 (1922).
39. ———, "Zur Physiologie der menschlichen Haut. IV, Über Zellpermeabilität und Erregung," *Pflüger's Arch. f. d. g. Physiol.*, 200: 278-284 (1923).
40. GILDEMEISTER, M., and ELLINGHAUS, J., "Zur Physiologie der menschlichen Haut. III, Über die Abhängigkeit des galvanischen Haut-Reflex von der Temperatur der Haut," *Pflüger's Arch. f. d. g. Physiol.*, 200: 262-277 (1923).
41. GODEFROY, J. C. L., "Het getransformeerde psychogalvanische verschijnsel," *Nederl. Tijdschr. v. Geneesk.*, 56: (1A, Part 1) 195-200 (1920).
42. ———, "The Psycho-electro-tacho-gram and Exophthalmic Goiter," *Psychiat. en Neurol. Bladen*, 26: 131-173 (1922).
43. ———, "The Variations in Shape of the Psycho-Electro-Tacho-gram," *Proc. VIII Internat. Congress. of Psychol.*, 8: 274-279 (1926).
44. GREGOR, A., "Beiträge zur Kenntnis des psychogalvanischen Phänomens," *Zeitsch. f. d. g. Neurol. u. Psychiat.*, 8: 393-412 (1912).
45. ———, "Die Hautelektrischen Erscheinungen in ihrer Beziehung zu Bewusstseinsprocessen," *Arch. f. g. Psychol.*, 27: 241-284 (1913).
46. GRÜNBAUM, A. A., "The Psychogalvanic Reflex Phenomenon and its Psycho-diagnostic Value," *Nederl. Tijdschr. v. Geneesk.*, 56 (1A, Part 1): 1044-1072 (1920).
47. JAMES, H. E. O., and THOULESS, R. H., "A Note on the Effect of Polarization in Psychogalvanic Reflex Experiments," *British Journal of Psychology* (Gen. Sect.), 17: 49-53 (1926-1927).

444 Diagnosing Personality and Conduct

48. JEFFRESS, L. A., "Galvanic Phenomena of the Skin," *Journal of Experimental Psychology*, 11: 130-144 (1928).
49. JONES, H. E., "Conditioned Psychogalvanic Responses in Infants," *Psychological Bulletin*, 25: 183-184 (1928).
50. JUNG, C. G., "On the Psychophysical Relation of the Association Experiment," *Journal of Abnormal Psychology*, 1: 247-255 (1907).
51. KAUFHOLD, R., "Über das elektrische Leitungsvermögen der überlebenden menschlichen Haut," *Arch. f. Anat. u. Physiol.*, 43: 189-216 (1919).
52. KNAUER, A., "Über den Einfluss von Ausdrucksbewegungen auf das elektrolytische Potential und die Leitungsfähigkeit der menschlichen Haut," *Klin. f. Psychol. u. Nerv. Krankheiten*, 3: 1-19 (1908).
53. KÜBLER, F., "Zur Technik der Darstellung des Psychogalvanischen Reflexphänomens," *Schweiz. med. Wchnschr.*, 24: 751-752 (1924).
54. KUHNÉ, V., "Résumé des Recherches Psycho-galvanometriques sur l'Émotivité de E. Abramowski," *Bull. de l'Inst. Gén. Physiol.*, 13: 157-191 (1913).
55. LANDIS, C., "Electrical Phenomena of the Body During Sleep," *American Journal of Physiology*, 81: 6-19 (1927).
56. LANDIS, C., and DE WICK, H. N., "The Electrical Phenomena of the Skin (Psycho-galvanic Reflex)," *Psychological Bulletin*, 26: 64-119 (1929).
57. LEVA, J., "Über einige körperliche Begleiterscheinungen psychischer Vorgänge mit besonderer Berücksichtigung des psychogalvanischen Reflexphänomens," *Münch. Med. Wochenschr.*, 60: 2386-2389 (1913).
58. LINDE, E., "Zur Frage vom psychischen Korrelate des Psychogalvanischen Reflexphänomens," *Proc. Eighth Internat. Congress of Psychol.*, 8: 351-352 (1928).
59. McDOWALL, R. J. S., and WELLS, H. M., "The Skin Constriction (Psycho-Galvanic) Reflex," *Nature*, 118: 644-645 (1926).
60. MORAVCSIK, M. H., "Experimente über des psychogalvanische Reflexphänomen," *Jour. f. Psychol. u. Neurol.*, 18: 186-199 (1912).
61. PEIPER, A., "Untersuchungen über den galvanischen Hautreflex (psychogalvanischen reflex) im Kindesalter," *Jahrb. f. d. Kinderheilk.*, 107: 140-150 (1924).
62. PETERSON, F., "The Galvanometer as a Measure of Emotions," *British Med. Journal*, 2: 804-806 (1907).
63. ———, "The Galvanometer in Psychology," *Journal of Abnormal Psychology*, 3: 43-45 (1908-1909).
64. PETERSON, F., and JUNG, C. B., "Psycho-physical Investiga-

Physiological Measures of the Emotions 445

- tions with Galvanometer and Pneumograph in Normal and Insane Individuals," *Brain*, 30: 153-218 (1907).
65. PETERSON, F., and SCRIPTURE, E. W., "Psycho-physical Investigations with the Galvanometer," *Journal of Nervous and Mental Diseases*, 36: 426-427 (1909).
 66. PHILIPPSON, M., and MENZERATH, P., "Sur l'Origine Musculaire du Phénomène Psycho-électrique," *Académie Royale de Belgique, Bull. d. l'Classe de Sciences*, 378-390 (1913).
 67. _____ and _____, "Analyse Physiologique des Phénomènes Psychoélectriques," *Arch. Internationales de Physiol.*, 15: 1-30 (1919).
 68. PIÉRON, H., "Sur les variations de la résistance du corps d'origine affective," *Compt. Rend. Soc. Biol.*, 77: 332-334 (1914).
 69. _____, "La Question du Mécanisme des Variations Psychogalvaniques Émotives," *Rev. de Psychiat.*, 16: 354-359 (1912).
 70. PRIDEAUX, E., "The Psychogalvanic Reflex: A Review," *Brain*, 43: 50-73 (1920).
 71. PRINCE, M., and PETERSON, F., "Experimentelle Untersuchungen über psychogalvanischen Reaktionen von mitbewussten (unterbewussten) Vorstellungen in einem Falls vielfacher Persönlichkeit," *Jour. f. Psychol. u. Neurol.*, 13: 249-262 (1908).
 72. RADECKI, W., "Recherches Experimentales sur les Phénomènes Psycho-électriques," *Arch. de Psychol.*, 11: 209-293 (1911).
 73. RICKSHER, C., and JUNG, C. G., "Further Investigations on the Galvanic Phenomena and Respiration in Normal and Insane Individuals," *Journal of Abnormal Psychology*, 2: 189-217 (1908).
 74. SCHILF, E., and MANDEN, I., "Zur Frage der Hemmungsinnervation der Schweissdrüsen," *Pflüger's Arch. f. d. g. Physiol.*, 196: 345-357 (1922).
 75. SIDIS, B., "The Galvanic Phenomenon," *Psychological Bulletin*, 7: 321-322 (1910).
 76. _____, "The Nature and Cause of the Galvanic Phenomena," *Journal of Abnormal Psychology*, 5: 69-74 (1910).
 77. SIDIS, B., and KALMUS, H. T., "A Study of Galvanometric Deflections due to Psycho-physical Processes," *Psychol. Review*, 15: 390-396 (1908), 16: 1-35 (1909).
 78. SIDIS, B., and NELSON, L., "The Nature and Causation of Galvanic Phenomena," *Psychological Review*, 17: 98-146 (1910).
 79. SMITH, W. W., "A Note on the Use of the Psychogalvanic Reflex," *British Journal of Psychology (Gen. Sect.)*, 12: 282-288 (1921).

80. SMITH, W. W., *The Measurement of Emotion* (Harcourt, Brace and Company, 1922).
81. SOMMER, R., "Zur Messung der motorischen Begleiterscheinungen psychischer Zustände," *Beiträge zur Psychiatr. Klinik* (Wien), 1: 143-164 (1902-1903).
82. STARCH, D., "Mental Processes and Concomitant Galvanometric Changes," *Psychological Review*, 17: 19-36 (1910).
83. STICKER, G., "Über Versuche einer objektiven Darstellung von Sensibilitätsstörungen," *Wiener Klinische Rundschau*, 11: 497-501, 514-518 (1897).
84. STROHL, A., and PORTES, F., "Sur les électrodes dites impolarisable," *Compt. Rend. Soc. Biol.*, 93: 310-312 (1925).
85. SUNG-SHENG, C., and LEHMANN, H., "Über den Einfluss verschiedener Spannungen auf den galvanische Hautreflex," *Pflüger's Arch. f. d. g. Physiol.*, 197: 437-440 (1922-1923).
86. SYZ, H. C., "Observations on the Unreliability of Subjective Reports of Emotional Reactions," *British Journal of Psychology* (Gen. Sect.), 17: 119-126 (1926).
87. ———, "Psychogalvanic Studies in Schizophrenia," *Arch. Neur. and Psychiatr.*, 16: 747-760 (1926).
88. ———, "Psycho-galvanic Studies on Sixty-four Medical Students," *British Journal of Psychology* (Gen. Sect.), 17: 54-69 (1926).
89. TARCHANOFF, J., "Über die galvanischen Erscheinungen an der Haut des Menschen bei Reizung der Sinnesorgane und bei verschiedenen Formen der psychischen Tätigkeit," *Pflüger's Arch. f. d. g. Physiol.*, 46: 46-55 (1890).
90. UHLENBRUCK, P., "Über den Autonomen Reflex," *Zeitschr. f. Biol.*, 81: 51-56 (1924).
91. VERAGUTH, O., "Das Psychogalvanische Reflex-Phänomen," *Monat. f. Psychiat. u. Neur.*, 21: 387-425 (1907).
92. ———, *Das Psychogalvanische Reflex-Phänomen* (Berlin, 1909).
93. VERAGUTH, O., and CLOTTA, G., "Klinische und experimentelle Beobachtungen an einem Fall von traumatische Läsion des rechten Stirnhirns," *Deutsch. Zeitschr. f. Nervenheilk.*, 32: 407-476 (1907).
94. VERAGUTH, O., and BRUNSCHWEILER, H., "Recherches sur le Phénomène Psychogalvanique dans Quelques Cas de Troubles Sensitifs par Blessures Cérébrales de Guerre," *Revue Neurologique*, 34: 151-162 (1918).
95. VIGOROUX, R., "Sur le rôle de la résistance électrique des tissue dans l'électro-diagnostic," *Gazette Médicale de Paris*, 1: 657-658 (1879).
96. ———, "Sur le Rôle de la Résistance Électrique des Tissus dans l'Électrodiagnostic," *Comp. Rend. de la Soc. de Biologie*, 31: 336-339 (1879).

Physiological Measures of the Emotions 447

97. VIGOROUX, R., "L'Électricité du Corps Humain," *Comp. Rend. Soc. Biol.*, 5: 138-142 (1888).
98. ———, "Sur la Résistance Électrique Considérée comme Signe Clinique," *Le Progrès Médical*, 7: 45-47, 86-88 (1888).
99. WALLER, A. D., *The Signs of Life* (London, 1903).
100. ———, "Galvanometric Measurement of "Emotive" Physiological Changes," *Pro. Roy. Soc.*, 90 B: 214-217 (1918).
101. ———, "The Emotive Response to Ordinary Stimulation, Real and Imaginary," *Lancet*, 194: 380-381 (1918).
102. ———, "Emotive Response of 73 Medical Students," *Lancet*, 194: 510-511 (1918).
103. ———, "The Influence of Drugs upon the Electrical Conductivity of the Palm of the Hand," *Proc. Roy. Soc.*, 91 B: 32-40 (1919-1920).
104. ———, "The Galvanometric Measurement of Human Emotion," *Nature*, 107: 183-186 (1921).
105. WASHBURN, M. F., HARDING, L., SIMONS, H., and TOMLINSON, D., "Further Experiments on Directed Recall as a Test of Cheerful and Depressed Temperaments," *American Journal of Psychology*, 36: 454-456 (1925).
106. WASHBURN, M. F., and PISEK, F., "Galvanometer Experiments with Revived Emotions as a Test of Emotional and Phlegmatic Temperaments," *American Journal of Psychology*, 36: 459-460 (1925).
107. WASHBURN, M. F., ROWLEY, J., and WINTER, G., "A Further Study of Revived Emotions as Related to Emotional and Calm Temperaments," *American Journal of Psychology*, 37: 280-283 (1926).
108. WECHSLER, D., "Sur la Technique du Réflexe Psychogalvanique (Présentation d'un Nouvel Appareil, le Photogalvanographe)," *Jour. de Psychol.*, 19: 172-177 (1922).
109. ———, "On the Specificity of Emotional Reactions," *American Journal of Psychology*, 36: 424-426 (1925).
110. ———, "The Measurement of Emotional Reactions: Researches on the Psycho-galvanic Reflex," *Archives of Psychology*, Vol. XII, no. 76 (1925).
111. WECHSLER, D., and JONES, H. E., "The Psychogalvanic Reflex in the Association Experiment," *Psychological Bulletin*, 24: 173-174 (1927).
112. WELLS, F. L., and FORBES, A., "On Certain Electrical Processes in the Human Body and Their Relation to Emotional Reactions," *Arch. of Psychol.*, Vol. II, no. 16 (1908-1911).
113. WELLS, H. M., "A Note on the Psychological Significance of the Psychogalvanic Reaction," *British Journal of Psychol.* (Gen. Sect.), 14: 300-308 (1923-1924).

448 Diagnosing Personality and Conduct

114. WELLS, H. M., "The Effect of the Circulation on the Electrical Resistance of the Skin of Man," *Quarterly Journal of Experimental Physiology*, 18: 33-44 (1927).
115. ZIMMERN, A., and LOGRE, B., "Le Réflexe Galvanopsychique," *Jour. de Radiol. et d'Électrol.*, 2: 610-619 (1916-1917).

Physiological Methods of Detecting Deception

116. ADLER, H. M., and LARSON, J. A., "Deception and Self-Deception," *Journal of Abnormal and Social Psychology*, 22: 364-371 (1928).
117. BENUSSI, V., "Die Atmungssymptome der Lüge," *Archiv. für die Gesamte Psychol.*, 31: 244-273 (1914).
118. BURTT, H. E., "The Inspiration-Expiration Ratio During Truth and Falsehood," *Journal of Experimental Psychology*, 4: 1-23 (1921).
119. LANDIS, C., and GULLETTE, R., "Studies of Emotional Reactions. III. Systolic Blood Pressure and Inspiration-Expiration Ratios," *Journal of Comparative Psychology*, 5: 221-253 (1925).
120. LANDIS, C., and WILEY, L. E., "Changes of Blood Pressure and Respiration During Deception," *Journal of Comparative Psychology*, 6: 1-19 (1926).
121. LANDIS, C., "Detecting the Deceiver," *Industrial Psychology*, 2: 244-249 (1927).
122. LARSON, J. A., "The Cardio-Pneumo-Psychogram and Its Use in the Study of the Emotions with Practical Applications," *Journal of Experimental Psychology*, 5: 323-329 (1922).
123. ———, "The Cardio-Pneumo-Psychogram in Deception," *Journal of Experimental Psychology*, 6: 420-465 (1923).
124. ———, "Modification of the Marston Deception Tests," *Journal of Criminal Law and Criminology*, 12: 390-399 (1921).
125. ———, "The Polygraph and Deception," *Welfare Magazine*, 18: 646-669 (1927).
126. ———, "Psychology in Criminal Investigation," in *Annals of the Teaching of Political and Social Science*, Vol. CXLVI (Nov., 1929), entitled *The Police and the Civic Problem*.
127. LOMBROSO, C., *Criminal Man* (G. P. Putnam's Sons, 1911), 322 pp.
128. MARSTON, W. M., "Psychological Possibilities in the Deception Tests," *Journal of Criminal Law and Criminology*, 11: 551-570 (1920-1921).
129. ———, "Systolic Blood Pressure Symptoms of De-

Physiological Measures of the Emotions 449

- ception," *Journal of Experimental Psychology*, 2: 117-163 (1917).
130. MARSTON, W. M., "Sex Characteristics of Systolic Blood Pressure," *Journal of Experimental Psychology*, 6: 387-419 (1923).

Acid Content of Saliva and Urine and Creatinine Content of Urine

131. FOLIN, O., "On the Determination of Creatinine and Creatine in Urine," *Journal of Biological Chemistry*, 17: 470 (1914).
132. HAMMETT, F. S., "The Practical Appreciation of Emotional Glucosuria," *Journal of American Medical Association*, 66: 1463 (1916).
133. HAWK, P. B., and BERGEIM, O., *Practical Physiological Chemistry*, 9th ed. (P. Blakiston's Sons and Co., 1926).
134. HENRY, G. W., and EBELLING, W. W., "Blood Calcium and Phosphorus in Personality Disorders," *Arch. of Neurol. and Psychiat.*, 16: 48-54 (1926).
135. HENRY, G. W., and MANGAM, E., "Blood in Personality Disorders," *Arch. of Neurol. and Psychiat.*, 13: 743-749 (1925).
136. RICH, G. J., "A Biochemical Approach to the Study of Personality," *Journal of Abnormal and Social Psychology*, 23: 158-175 (1928).
137. ———, "The Reaction of Human Mixed Saliva," *Quarterly Journal of Experimental Physiology*, 17: 53-56 (1927).
138. ———, "Some Biochemical Aspects of Personality Traits," *Psychological Bulletin*, 24: 187-188 (1927).
139. STARR, H. E., "The Hydrogen Ion Concentration of the Mixed Saliva as an Index of Fatigue and of Emotional Excitation and Applied to a Study of the Metabolic Etiology of Stammering," *American Journal of Psychology*, 33: 415-417 (1922).
140. THOMAS, W. I., and THOMAS, D. S., *The Child in America* (Alfred A. Knopf, 1928), 583 pp.

Chapter XII

INTERVIEWING

INTERVIEWING is a method of securing data by a face-to-face consultation or conference in which a person tells an interviewer his story or version of a set of facts or answers questions concerning whatever problem or inquiry is at hand. The interview may employ observation and rating as techniques, and it bears close resemblance to the questionnaire in many respects. While in one sense interviewing has greater flexibility than any one of these methods, in that no schedule need be rigidly adhered to, in another way it lacks flexibility, in that the mental processes of another person are often intricate and intractable and attain a certain momentum, making it difficult to hold the conversation to the main topic. Since in the past some of our more exact testing methods have been developed from the interview, there is every reason to believe that interviewing is now in process of elaborating through experience more accurate methods for investigating certain phases of conduct than are now available. In the interview the *measurement* point of view is lacking, and the interviewer must make an interpretation from running narrative or isolated answers to questions. Though this is the ordinary common-sense way of sizing up a situation, it can hardly be thought of as an exact method.

Types of Interviews

Three main types of interviews should be distinguished. The *diagnostic* interview seeks to discover facts concerning the life history of the informant and to extract from his narrative his opinions, attitudes, and personal experiences. In the diagnostic interview interest is concentrated on the individual being interviewed. In the *research* interview interest does not lie in the person being interviewed, except as he can contribute data helping to solve a particular problem. A good example of this type of interview is furnished by the work of Charters (17) in making job analyses, as for example, when secretaries were interviewed to discover the

operations which they performed, with attention centered on the jobs and not on the person being interviewed. The third type of interview includes *treatment*. The social case worker is interested not only in obtaining facts regarding her client, she wishes also to persuade him to change somewhat his family situation or other adjustment. Or the college dean hopes to effect through the interview some improvement in the standing of the student before him. The *diagnostic* and *treatment* interview are often intermingled, and it is difficult to tell where one begins and the other leaves off. In the present discussion our concern is strictly with the diagnostic phases of the interview, and the interviewer's endeavors to influence his client must be ignored.

Another difficulty in discussing the interview is that interviews conducted by different persons differ in purpose and in technique. Interviews are conducted by physicians, lawyers, priests, journalists, detectives, social workers, psychiatrists, psychoanalysts, deans, research workers, employment managers, anthropologists, sociologists, and others. The physician wishes a clear and accurate statement of disturbing symptoms, the priest wants a confession, the newspaper reporter wants a racy story or an unusual and significant comment, the social worker is anxious to discover social relationships, employment managers are on the lookout for the aptitudes and character traits that make a man a valuable employee. The sociologist is particularly interested in the environment, whereas the psychiatrist allows his interest to become absorbed by the mental mechanisms uncovered. Each purpose requires its own technique, and it is difficult to describe them together under the single heading "interviewing."

Advantages and Disadvantages of the Interview

The interview has often been contrasted with the *questionnaire*. Each resembles the other. In both methods questions are asked and the answers are to be taken at their face value as facts, or used to show important trends in the individual.

Charters, who has made effective use of the interview in determining curriculum objectives, says of this method in his *Curriculum Construction** (19, p. 134):

* From Charters, W. W., *Curriculum Construction*. By permission of The Macmillan Company, publishers.

"In all cases the oral questionnaire is preferable to the written form. This means that when the teacher becomes an interviewer and asks the questions orally, he will obtain more reliable answers. He can clear up misconceptions of his meaning, and can supplement his questions by others which will elicit more definite answers. The chief objection to the oral questionnaire is the labor required of the interviewer; but this is amply compensated for by the greater trustworthiness of the information, the comparative ease with which the answerer is able to provide it, and the greater amount of time he will spend upon it. Answerers who dislike to write, and would spend very little time on written answers, are glad to devote considerably more time to an oral interview."

Whether or not the examiner will follow up leads during an interview, in order to obtain greater explicitness than is possible in the questionnaire, depends on his versatility and his desire to complete an adequate piece of investigation. Meltzer (39) claims that the interview yields more meaningful and revealing responses than other methods of inquiry. He lists as advantages of the interview method that the responses are more natural, less artificial, more informal, and more spontaneous. The interview gives the examiner opportunity to investigate further generalities, obscure points, or evasions. It permits the examiner not only to collect facts, opinions, attitudes, likes, and dislikes, but to find out the "why" of these responses. Cavan points out that through the interview it is possible to trace development more effectively than with the questionnaire.

But the interview has its disadvantages. Suggestion has as much force in the oral interview as in the questionnaire, if not more. Witness responses to a doctor's request for symptoms. If you are an "ailing" sort of person, you enlarge upon insignificant symptoms; if you are a stoic, you conceal what may be important symptoms. Likewise in any interview a person's judgment will be pulled or warped not only by the questions and their form, but by expression, intonation, gesture, and the like. "You only study a half-hour a day?" said with an air of surprise brings the defense response, "Why, usually I spend more time than that, but sometimes I can get it in half an hour, and then I have my music to practice, you know."

The interview, in addition, is much more expensive in time and cannot reach as large a group of informants as the questionnaire with an equal expenditure of time and effort,

Finally, because the interview as free response cannot be "scored," it cannot be quantitatively treated and interpreted. This denies to the interview amenability to accurate interpretation that can be accorded the data from a carefully assembled questionnaire.

Davis (23) reports a comparison of the questionnaire and interview methods by McCall that has suggestive implications. The replies of fifty women interviewed as to their sex life were compared with the questionnaire, replies of fifty women of as nearly equal age, education, and mental status as possible. Unfortunately the differences reported in this study were not treated statistically for reliability, and there lurks a suspicion that many of them result from mere chance. The main findings are as follows:

1. The questionnaire method is more accurate, judged on the basis of eliciting a larger number of admissions of sex practices among the group. Since these practices are socially disapproved, we are tempted to generalize from this finding that an anonymous questionnaire is more reliable than an interview in revealing disapproved practices.

2. The interview method is more reliable in that it yields fewer inconsistent replies. This may be due to ill-defined questions in the questionnaire or to the fact that, if necessary, questions in the interview can be elucidated further.

3. The interview method did not produce appreciably more answers with *definite* information (i. e., replies with entire frankness and candidness).

4. Slightly more information relative to childhood impressions and feelings was secured by the questionnaire.

5. While anonymity of the questionnaire method may contribute a distinct advantage over the interview method if a follow-up of individual cases is desired, the very anonymity of the former method renders a follow-up investigation impossible, except of course where additional general questions are addressed to the whole group.

6. The interview method secured 34 per cent replies of those solicited against 32.2 per cent for the questionnaire—a negligible difference.

7. That the interviewer contributes anything through his judgment of the personality of the subject or through detection of dishonesty is doubtful.

When we consider the greater ease with which the questionnaire

method may be employed and that it may be depended upon to yield fully as complete and accurate and frank results as the interview, the results of this study seem to indicate that the questionnaire method is preferable.

Standardization of the Interview

The matter of *standardization* is a point of issue concerning the interview and one that has divided interviewers into two camps. On the one hand there are those who plead for more flexibility in the interview. The interviewer is admonished to adapt the interview to the individual, so far as possible becoming all things to all people. He is told when conducting an interview, to "do in Rome as the Romans do" and to adopt the language, mannerisms, dress, and attitudes of those he is questioning. He is told to follow no fixed schedule of questions but to endeavor to get his subject to converse naturally and to use questions only to follow up promising leads. He should analyze the story as it proceeds and be versatile in redirecting the thread of thought wherever clues are most promising. The object of this type of interview is to draw the informant out and obtain an unconstrained account of his prejudices, motives, drives, passions, etc.

On the other hand there are those who advise planning the interview carefully ahead of time. The questions should be framed, printed in a schedule, perhaps rehearsed. Hamilton (27), in his study of marriage relations, was so fearful lest the intonation of his voice should suggest an answer, that he prepared all of his questions on typewritten cards and handed them to his subjects one at a time. The responses were as free and untrammelled as possible. Others have gone still further and have not only prepared the questions carefully beforehand, but have attempted to standardize the selection and formulation of the questions much as is done in test construction. Snedden has prepared a disguised intelligence test in the form of an interview, while O'Rourke (46) has prepared problems in interview form for testing the judgment of applicants for the position of prohibition agents. Used in this way the interview approaches the questionnaire or test in degree of standardization.

Those who would make the interview flexible are keenly aware of the difficulty of getting persons to express themselves freely.

Those who would standardize the interview are most conscious of the vicious influences of suggestion in biasing the results, of the danger of laying too great stress on unprofitable clues or the answers to insignificant questions. Future developments are certain to lead to increased experimentation with methods of standardizing the interview. Since our experience with tests has shown how extremely important to the results the manifold variable factors in the situation are, it is imperative in the interests of increased accuracy in the results of the interview that there should be a more thorough try-out of uniform methods. On the other hand, we should avail ourselves of every device and technique for enlisting the rapport of the subject, for breaking down resistance, and for encouraging free expression.

The Interviewer

Qualifications. Since the interview is so subjective a process and one in which personal qualities play so large a part, the personality of the interviewer is an important factor. In the discussion of tests, questionnaires, etc., our main concern was with the accuracy of the data. In the interview the main consideration seems to be getting any data at all. An interviewer is rated primarily for his success in drawing subjects out and getting them to talk freely about themselves.

The interviewer should first of all be thoroughly familiar with the field of investigation. If he is a school counselor or college dean, the interviewer should know all about the institution—its curriculum, its traditions, and its clientele. If he is a vocational counselor, the interviewer should know all about industrial and business opportunities, working conditions, type of job available, etc. If he is a social worker, he should have adequate knowledge of the social environment of the home—the churches, schools, clubs, stores, and places of amusement.

Charters (18, p. 285) wants his interviewers to be "logical-minded." What does this mean except that they should be generally intelligent and keen-witted to sense the forces at play in the life of the person whose story is being unfolded? Charters also wants his interviewers able to "dig in," by which he means able to follow up promising clues to elaborate some point on which the subject fails to go into detail.

Griffitts (26) speaks of a knowledge of psychology as being essential to the equipment of the interviewer. Much depends on the psychology, however, and certain forms of academic psychology are little more than useless. A detailed knowledge of mental mechanisms, however—the rationalizations, attitudes, and motives that guide men's behavior—should be of aid in interpretation.

The interviewer should be well adjusted himself. The psychiatrists tell us that one cannot be a successful interpreter of the mental mechanisms of others until he has seen first to his own adjustments. The interviewer should be poised, emotionally stable, not prone to show undue excitement, and he should understand himself. Some say that one must actually have lived through various forms of experience before he is able to clearly recognize and diagnose them in others.

The interviewer should command the respect of his informant without awing him. His position should be one of prestige in the eyes of his subject, but he himself should be friendly and cordial. His personal appearance must be acceptable. There must be nothing in the appearance or approach of the interviewer that will tend to offend or disgust the other person. He should be plainly, neatly dressed, not conspicuously nor untidily. He should be as educated and as refined as those with whom he is to deal, but if possible not more so. On some occasions he should show considerable *savoir faire*; on other occasions he should adopt the language, gestures, dress, and bearing of the person he is to interview.

Sincerity and *sympathetic understanding* are important characteristics for the interviewer to possess. He must be genuinely interested in the problems and troubles of other persons, and be ready to listen attentively to a long-drawn-out tale of woe. He should be able to see the other's viewpoint and to share the other's hopes and fears whole-heartedly, without a tendency to ridicule or to criticize them. Kindness and tolerance rather than coldness, austerity, and superciliousness should characterize the interviewer. He should neither approve nor condemn the errors and lapses of his subject nor should he exhibit surprise nor maudlin sympathy, but genuine whole-hearted understanding. He should not betray the slightest trace of surprise or shock at the disclosures made to him, but receive everything in an impersonal.

detached, unsentimental manner. The person interviewed must be convinced that he is talking to a friend who will not betray him and who is primarily interested in his welfare.

The interviewer must be courteous and respectful, displaying the deference which is appropriate to the occasion. Seldom should he contradict or dispute another's point of view, except where this is done purposely to make a man take sides and express himself vigorously. The interviewer should be a person of vigor, one who will command the attention of the informant. He should have the warmth and heartiness about him that breeds a sense of confidence and security.

Finally, the interviewer should have a sense of humor. Often in an interview tensions and inhibitions will arise which block the free development of the story. The interviewer must be on the watch for these emotional blocks, and stand ready to ward them off. A sense of humor sometimes will help one in smoothing over a situation which is developing uncomfortably.

Training. Although many of these characteristics are so fundamental that the interviewer must be selected for them rather than trained in them, there are certain matters of technique in which training should be of distinct advantage. Schools for training psychiatric social workers give definite instruction in the art of interviewing. There are definite techniques in making the approach, gaining rapport, and breaking resistance. The groups which believe in a more complete standardization of the interview process gives instruction in the precise choice and formulation of questions, in the recording of answers, and in the interpretation of the answers. Accurate studies of the interview technique have yet to be made, but we may surmise that training the interviewer will enhance greatly the reliability and significance of the interview.

Voluntary or Involuntary Interviewing

Should the interview be voluntary, or may it sometimes be involuntary? All depends on the situation. The applicant for a job, the confessor before the priest, and the sick man before the physician seek the interview. On the other hand, testimony in a court of law is often decidedly involuntary, as when the witness is activated not by the primary wish to answer any ques-

458 Diagnosing Personality and Conduct

tion that may be put to him, but by a desire to clear himself or others of accusation. In school the interview is sometimes optional, as when a student seeks advice or guidance; at other times it is involuntary, as when a boy is sent to the principal's office for misbehavior.

When the interview is voluntary and initial resistance is lacking, one may expect the answers to be on the whole truthful and complete. Even in the voluntary interview, however, the subject may either intentionally or unintentionally withhold some of his story, or color it so as to show himself in the most favorable light. But it is in the involuntary interview that all of the arts of the interviewer must be used. At every step some sort of appeal to the self-interest of the subject must be employed. To persuade him to consent to the interview in the first place he must be promised that it will help him attain the position, success, or strength that he craves, or understand himself better, or avoid failure or incompetence. Sometimes in the research type of interview it is possible to obtain cooperation by referring to the subject's position of prominence or by remarking about his expertness. "We are canvassing the opinions of some of the outstanding men in the freshmen class" is almost certain to break down resistance and reserve. If the name of some person or institution can be used to add authority to the interview, this may help to overcome initial resistance. The old trick of the book agent who sells the first book in a town to the minister of the local church need not be employed, but any just cause for an interview can usually acquire institutional backing.

Preparation for the Interview

The preview. As a preparation for the interview one should try to find out everything possible about the subject. A newspaper reporter interviewing a celebrity should have read his biographical sketch in *Who's Who* and should know all about his recent exploits. The school counselor before interviewing a pupil should consult the school records and be familiar with such facts as his age, grade, teachers, marks, participation in extra-curricular activities, intelligence test score, physical examination record, and whatever else the recorded evidence may yield. A knowledge of his reputation with his various teachers, his out-

standing behavior characteristics, hobbies, interests, and noteworthy achievements would be of help. The social worker should be thoroughly conversant with the past history of his client's case, from the occasion of original reference through all succeeding interviews and recommendations. Armed with such preliminary information, the interviewer will be in a position to make an effective opening, to gain rapport, and to procure the desired additional data most readily.

Making initial contacts and appointments. Gray and Monroe (25), in interviewing adults on their reading interests and habits, paid particular attention to establishing initial contacts. A letter was sent to make the first contact and this was followed by a telephone call to make the appointment. A definite time for the interview should be set and this time should be rigidly adhered to by the interviewer. In cases where the initial appointment cannot be arranged by telephone, a letter of introduction may be used.

Conditions of the interview. The time and place of the interview are important factors in determining its success. Ample time should be provided for an interview, since the appearance of leisure to talk over problems to any desired length is essential. Those who see the interview as an informal way of getting facts recommend that the interview be held wherever and whenever it is most convenient to the subject. The reporter attempts to catch his subject on leaving the boat or boarding his train or in his hotel. Time and place are made to fit the convenience of the person being interviewed. The social worker finds that it is sometimes helpful to hold the interview in the home, not only for convenience to the subject, but also because one then gets a clear picture of the environmental background of the case.

On the other hand, strong arguments have been proposed for the office interview. In the office the *privacy* and *freedom from distracting influences* can be had which it is usually not possible to get at home. For the type of interview where inhibitions must be broken down, an atmosphere of security and time to talk over problems is requisite. The interviewer must breed assurance that he is the sort of person in whom one may confide in safety, and that he has a genuine interest in the welfare of his subject. Balanced against the privacy of the office is the sense

460 Diagnosing Personality and Conduct

of strangeness which in some cases may become an inhibiting influence.

If one is to hold interviews in his office, some attention should be paid to its atmosphere. The waiting room should be pleasant and comfortable, with a living-room atmosphere that is at the same time professional. Comfortable chairs and interesting magazines must be provided. The office itself should be neat and businesslike, the fewer distracting objects about the better. Probably in the interests of privacy, secretaries or stenographers should be out of sight. Subjects should be made as comfortable as possible, perhaps even reclining so that there may be comfortable relaxation.

In general we may conclude that if we are primarily interested in the sociological background of the subject, he should be interviewed at home or in some other characteristic haunt; but if the purpose of the interview is to get the subject to disclose his own attitudes and problems, perhaps the office with its privacy and security is the best place.

Steps and Procedures in Interviewing

Those who have analyzed interviews find that they shape up much like any other story. In form they bear resemblance to the novel or the drama. Four separate stages have been recognized: the introduction, rising action, dénouement, and conclusion. Sometimes one or another section is lengthened out or telescoped, but they all appear unmistakably. Each step of the interview has its own techniques, and a discussion of them will illustrate the procedure.

Introduction. The introduction may be brief or long according to circumstances. If the interview is voluntary and the problem clear-cut, it may proceed at once to the point at issue. Usually, however, there must be a brief period of getting acquainted. The proffer of a cigarette, an invitation to lunch, a ride in the car may serve to break down reserve. In the office interview the interviewer may meet his client with a hearty handshake. The usual bromidic phrases about the weather, or remarks about the latest news or rumors always serve as conversation starters.

Often one needs at the outset to describe briefly but honestly the purpose of the interview. Students may come to a personal

conference with a counselor, holding an entirely wrong notion of its purpose. The first requisite to put the student at ease, allay apprehension, or break down diffidence is a simple statement of the purpose of the interview. At this time one should stress the advantages that can come to the interviewee from the conference, or, if it is a research interview, the contribution he can make to the study under way. The subject must be persuaded that the interview is to be worth while, that the interviewer is competent, and that he can handle the problem professionally. The interviewer should do everything possible to relieve the subject of fear or embarrassment and to assure him of his own genuine desire to be helpful and his intention to give a square deal.

In the case study interview, where one wishes to cover the topics of a prepared schedule, it is wise to open the conversation on some topic or hobby of special interest to the person interviewed. Most persons are quite ready to talk about themselves. Maulsby (38) suggests that such questions as, "When are you going to get married?" or "When are you going to set up business for yourself?" never fail to unleash the tongues of the person questioned because they tap the deepest aspirations and longings of most persons. Similarly for school-children, "Are you going to camp this summer?" or "Who is going to win the game this week?" or "What kind of marks are you getting this term?" break down shyness. The latter is a valuable lead, especially if the marks have in the past been good. In general the interview should open on topics dealing with successes and triumphs rather than on those connected with failure.

Sturtevant and Hayes (63) suggest that the first question in an interview or any point involving personal difficulty should relate to another individual rather than to the person himself. For instance they advise asking, "Does your father scold you about spending money?" rather than "Do you spend more money than you ought?" This flank attack avoids the direct semblance of criticism, and helps to steer the subject away from rationalization.

Securing rapport. Often, if there is suspicion or hostility on the part of the subject, definite techniques must be employed to disarm him, gain his confidence, and secure rapport. One must bring himself down to the subject's level, use slang or colloquial

language where possible. But one must not "talk down" to a child. The interviewer should convince the subject that he is an insider, worthy of confidence, rather than an outsider. This means that the interviewer must have had considerable contact with the social group represented by the subject.

If possible the interviewer should tie himself up in some way with the subject's past experience. A reference to common friends often makes an excellent bridge with which to span the gap. The discovery that both have lived in the same village, visited the same city, taken the same trip, enjoyed the same show, or attended the same school will serve as an effective means of gaining rapport.

If possible the subject should be made to feel that he is leading the interview. This can be done by giving him free rein in expression and asking questions which follow up leads already opened.

Another suggestion by Brisley (10) is that something should be done to relieve the timidity or tension or embarrassment at the beginning. The social worker may offer some immediate measure of relief for the present need, such as new clothes for the children, or food for a short time. The school counselor may promise to intercede with a teacher or the school principal. Once the tension has been removed from the immediate situation, the subject is freed for a more rational appraisal of the total situation.

Rising action. When by means of the introduction a satisfactory degree of rapport has been established, one should come directly to the point at issue. At this point a question of technique arises as to whether it is better to allow the subject to tell his narrative with as little coaching or prodding from the interviewer as possible, or to begin by asking a series of questions. Experimental work by Cady (15) and earlier workers has demonstrated that more ideas are expressed, and expressed with less error, when the free narrative comes first and is followed by more detailed questions than by any other combination. Consequently it is best not to grill, coerce, give advice, or show authority at the opening of the interview, but to wait until the subject is ready to tell his own story. The interview is not a cross-examination, but a matter of coöperation.

A kindred issue upon which authorities differ concerns broach-

ing the immediate problems first or letting them wait until something of the history of the subject has been revealed. Mowrer believes that the immediate problems should not be discussed until the subject has had an opportunity to tell his life story, or story in general, a method which postpones coming to the disagreeable issues until a feeling of confidence and familiarity has been established. On the other hand, under certain conditions it may be well to come to the point at issue directly so that testimony from the life history can bear upon it.

It is unanimously agreed that the subject should be allowed to express himself freely, and even to ramble if need be, at the opening of the interview. Most persons in trouble have a certain amount of pent-up emotional irritation that has to be discharged before they can think calmly and rationally on their problems. Let them blow off steam and tell about the injustices that they have been subjected to, the accidents and misfortunes that are theirs, the jealousies, spites, and revenges that they are harboring. During this stage of the interview the interviewer must show much patience and forbearance. He must make every effort to really see his subject's point of view and to agree with him so far as possible at every stage. To contradict, to argue, to dispute at this point would be fatal. The subject should be allowed to "get it off his chest" without interruptions. At this stage plenty of time should be allowed.

Certain definite techniques for encouraging free expression have been analyzed by Palmer (47). First of all are the introductory remarks or questions which indicate a genuine interest on the part of the interviewer. Then there are the gestures, nods of the head, smiles, encouraging remarks such as "Uh-huh," "Isn't that interesting?" or "Isn't that strange?" Facial expressions which reflect sympathy with the emotions in the unfolding tale help the interviewee to warm up to his story. Nearly every one responds to an appreciative audience. The reticent individual who finds it difficult to disclose his personal secrets should not be told to "go on" but should be asked, "Is there anything else?" or "What about ——?" Definite leads into the unknown are more potent than prods from behind.

Nothing is more important in the interviewer's technique than being open and frank himself. There should be no occasion for an interviewer saying "I can't take that up with you just now"

or using some device for turning the topic. The interviewer cannot expect his subject to be open and frank unless he is so himself. If his genuine wish is to help his client, there should be nothing to be hidden or withheld. The person being interviewed should have access to everything that he desires to know which will help him understand the situation. The interviewer should particularly avoid being put on the defensive himself, as this would defeat the whole purpose of the interview.

Motivation. One who undertakes interviewing must have a thorough knowledge of human motivation and skill in adapting this knowledge to particular situations. He must, of course, know in general what types of appeals are effective, but since not all persons respond to the same appeals, he must know also what these different appeals are usually associated with. It all comes down finally to a matter of knowing the individual concerned and what forces he is responsive to.

As an illustration, take the appeal to prejudice. Some prejudices are nearly universal, but others, more specific, are related to race, nationality, religion, sex, occupational group, economic level, or political party. The interviewer, to use prejudice as a motivating force, must know the common prejudices, what they are associated with, and whether these factors apply to the individual being questioned. The appeal which will arouse the prejudices of a labor-union member will have just the opposite effect on a wealthy member of the Republican party.

Appeals may be made to ambition, pride, ideals, weaknesses, foibles, scruples, desires, tastes, esthetic sense, sentiment, sense of humor, sense of justice, and altruism, and these are but a few of the countless factors which at one time or another may be used for motivation. The appeal will be by that much more effective, however, the more sincere and genuine the basis on which it is made.

Techniques for lessening tension and meeting resistance. Rannels (53) has analyzed the situations which cause tension and resistance on the part of the person being interviewed. She lists five groups of causes: (a) Causes arising from environmental feelings. The setting or situation has already been referred to as a factor in making for free expression or leading to inhibition. (b) Causes arising from differences between the social worker and the interviewee. (c) Causes arising from the interviewee's

intellectual or emotional reaction to the particular subject or situation under consideration. (d) Causes arising from the habitual reaction pattern of the interviewee. Rademacher (52) has given a very penetrating analysis of some of these habitual reaction patterns. His whole discussion should be read, but the following quotation illustrates his main thesis (52, pp. 82, 83):

... "in cases where a child is spontaneous, enthusiastic, and talkative, analysis of the material shows, after all, very little of an actual conflict situation revealed verbally. Often the very verbalism of the patient, while it makes for interesting and lively staff-discussion material, defeats the psychiatrist's attempt to unearth a discussion of subjective attitudes. There are many reasons why this is so. Of importance is the fact that usually such material is painful for the patient to discuss. Only certain types of children derive pleasure from a discussion of their inadequacies or inferiorities, and in such cases these deficiencies are not important sources of conflict. Again, these verbalistic children are so simply because they have learned the value of this method in gaining their own ends. They have discovered the value of verbalizing or becoming circumstantial as a method of evading unpleasant situations or of getting them out of assuming responsibility or putting forth effort toward any accomplishment. In other children verbalism is the essence of the conflict situation itself, in that these children often find adjustment to children of their own age difficult because of their close contact with their parents and their parents' friends. The child knows how to impress adults and attempts to do this with the psychiatrist in his effort to avoid discussion of the unpleasant relationships that actually are his with other children.

"There are many reasons why only a minimal amount of actual subjective feeling is obtained from a child in an initial contact. A child who is afraid in new situations is from the start under an emotional strain which does not allow him to converse freely. His responses become stereotyped and his attention inadequate. A child who is apprehensive as a result of some recent misdemeanor will also react to that fear. Usually he will attempt to rationalize his conduct, to show his innocence in the matter, or to place the blame elsewhere. He will spend much time attempting to impress the examiner with his good intentions and his high ideals, and will not face issues that involve his personal responsibility. The child who comes from a home in which discipline has been repressive is often so suppressed that he is actually afraid to express his feelings on any subject. Often, too, such a child, as a result of his suppression, has no feeling of security in any of his relationships. The home, which may be at the basis of his difficulty, offers the only security that he has, so that

he will not readily commit himself in any way that may affect this tie. Children who have been accustomed to overprotective attitudes in the home have difficulty in facing situations that require personal effort. As a result, their behavior becomes of a compensatory or extravagant nature with the aim of bringing attention to them easily. Otherwise, because of a personality molded by oversolicitous care, they are ridiculed by other children. They do not like to face the idea that they are not well suited to meet outside responsibilities or that they are dependent on their parents. At the same time, this very situation is an agreeable one to them and naturally they do not wish to give it up. The responses of these children are often what might be termed 'the proper thing to say' rather than an expression of their actual feelings. Problem children from homes in which discipline is inconsistent and highly tinged with emotion realize that notwithstanding the nagging, scolding or whippings they occasionally receive, such a situation works to their advantage in the sense that they are able to have their own way. Therefore, their responses are quite likely to be conventional or stereotyped ones."

(e) The fifth group mentioned by Rannells consists of causes arising from the uncertainty of response (53, p. 92):

"Fear of disfavor in the eyes of the social worker or the community; the danger of ostracism from a particular social, religious, or professional group if the adherence to certain non-conformists' views is discovered; the apprehension of inadvertently betraying another's confidence; the desire to avoid legal entanglements which might result from the disclosure of certain facts—all of these may act as an impediment to the setting up of a natural and friendly relationship."

Definite techniques are available for lessening tensions and meeting resistance. One device is simulated agreement. The writer discovered how effective this device could be in dealing with a certain student when a point of disagreement had been under discussion and each held to his point of view stubbornly. Finally I found an error in my position and yielded. To my surprise the student immediately yielded not only the main point at issue but several other points which he had consistently maintained. My confession of error broke down his resistance at many points.

A second device is that of minimizing the seriousness of the client's position. This can be done in such various ways as quoting the many favorable factors, the assets of the situation, or mentioning other people who are having the same difficulties,

or citing statistics to show the relative unimportance of the point at issue. Frequently persons get upset because the unfavorable factors in the situation loom so large that they are seen quite out of proper perspective. To recall to them the insignificance of the things that give them fear and worry as compared with the neglected favorable factors in the situation is often reassuring. But enlightenment of this kind must be concrete and simple enough to influence comprehension. Most persons can be influenced more effectively through anecdotes than through statistics, even though the anecdote really proves nothing.

Another device is that of analyzing a general statement into specific parts. The school pupil may believe himself to be a failure when analysis shows that his difficulties lie in only one subject. Quite frequently a specific experience is made the basis of a generalization which is then interpreted as rendering the whole situation irretrievable or dismal.

The converse technique is to introduce a general statement to displace attention upon specific details. The boy who resents punishment for a misdeed should be told that his teacher is fair and just, or the football player who misses the punt needs to be reassured that he rates as a star player.

An entirely different method for lessening tension employs physical contact between interviewer and his subject. The hand on the shoulder breeds confidence, or the touch of a hand furnishes just that slight suggestion of intimacy that conveys assurance and allays suspicion.

Other devices for meeting resistance have been alluded to before. The skilled interviewer betrays no shock or surprise at any admission. The boy who confesses going with bad companions, cheating, theft, or masturbation should encounter in the interviewer no critical or reproachful attitude, but a quiet acceptance of these admissions as so much evidence. The interviewer should evince no surprise or discouragement when his client admits divorce or an illegitimate child.

The emotional reactions of the client—anger, weeping, sullenness, shouting, berating, and the like—must be ignored so far as possible by the interviewer. Sometimes these are reactions to the situations under discussion, at other times they are simulated attempts to influence the interviewer. At all events the interviewer should not be swayed by them. He should not talk back angrily,

nor show undue sympathy to weeping, nor should he coax or threaten.

Evasions of the interviewee, however slight, must be noted. Topics which are evaded should not be pressed, at least by a frontal attack, but sometimes it is possible to come around to them easily through collateral topics. At other times it is wise to postpone their discussion until a later date when other facts have been uncovered which make the point at issue more easily approached.

Still other devices for breaking resistance come under the headings "jollyng," "flattery," "humor," and the like. It seems too bad that such cheap devices as these should have to be considered in listing techniques for dealing with other people. But human beings have their susceptibilities and their failings, and these represent avenues of approach that are effective. Our philosophy often must admit that the end justifies the means, particularly when the end is a worthy one and the means are not in themselves harmful.

Bogardus (7) has listed several mechanisms of mental release. First he mentions the *naïve* type of release, exhibited in persons who are only too glad of an opportunity to converse freely and willingly, to talk shop, to relate an incident or to make a confession. The second sort of release mechanism is called *egotistical*, because an appeal to vanity or pride will unlock the doors of reticence. A third mechanism is the *confession*, which by releasing pent-up tensions gives relief. A fourth mechanism of release is termed *scientific*, to describe the attitude of the person whose deposition is "for the sake of truth and science," who takes an objective viewpoint concerning himself, who scorns much of the conventional code of judgments of "right" and "wrong," and who is even willing to "sacrifice" himself "for the sake of science." The *sophisticated* type of mechanism is shown by those whose experience makes them inaccessible to the ordinary appeals—toughened individuals such as lawyers and physicians who know too much to "spill" it to the inquisitor ingenuously.

Breaking down defense mechanisms. Where there are definite mechanisms at work which cause the subject to defend a position more or less stubbornly, resort to extreme measures capable of stirring powerful motives may be necessary to dislodge him from an intrenched position. Some of these measures are taken

from salesmanship, others from legal practice. Although, in general, measures which approximate coercion or force should not be employed, since the interview should be a coöperative enterprise, they are sometimes necessary. The first rule is to avoid so far as possible putting another person on the defensive. If the interview is properly opened and rapport obtained, there will seldom be provocation for a person to adopt a defense mechanism.

One useful bit of strategy is to describe the ultimate outcome of things as they are going. To a pupil in school one may hold up the picture of failure at the end of the year, expulsion from school, failure to get into college, as being the outcome of his present behavior. Though this may to some extent involve fear, the endeavor should be to have the subject project himself by anticipation into the future without emotion. It is most important that he see his present situation in the light of its eventual consequences.

A second device is "abusing for defense." It is a well-known fact that most persons cannot bear to hear criticized or berated certain persons or things which they themselves will at times criticize or abuse or perhaps neglect. If the interviewer systematically proceeds to criticize a pupil's parents, or his teacher or school, in all probability the pupil will rush to their defense, forced to find commendable features that no other technique could get him to recognize or admit. This is a dangerous device for the interviewer to play with, but is often most effective.

Other related devices are listed by Salsberry (57) under the headings "puncture," "rushing," "swaying by oratory," "taking client off his guard," "using acquired information," "putting cards on the table," "closing into a corner," "instilling fear," "negation," etc. Certain of these overlap, others are names of devices to be used only in unusual situations. Suppose a subject's story has all the earmarks of plausibility but contains one point markedly inconsistent with other facts. By decisively singling this out the subject's whole story can be "punctured." Harsh measures for putting the subject in a corner, instilling fear, taking him off his guard, and the like should be avoided when possible.

Keeping the interview to the main issue. In the interview another set of techniques is necessary to ensure keeping to the main issue. At first the subject may be allowed to ramble until he has exhausted the high pressure of his pent-up emotions

through the safety-valve of free expression. Sometimes as he does this a canny subject may set out gradually to veer away from the issue by planned digressions. In any case the time comes in every interview when the main issue should no longer be dodged.

One of the simplest and most direct methods of keeping the interview to the topic is by the use of questions. When the narrative is finished, the interviewer proceeds to fill in the gaps and uncover the obscurities by direct questions. If the main issue is still sensitive, then the probing may be so indirect and along collateral lines that it gradually breaks through the outer defenses by disclosure after disclosure until the resistance to the main issue is weakened and finally overcome.

Salsberry (57) mentions other devices, some of which are mere repetitions of techniques used for other purposes. Failure to answer digressive questions put by the subject may help to maintain the line of advance. It takes two to keep a conversation alive, and the interviewer can cause some topics to expire by merely failing to respond to them. "Sharing personal experiences of the same nature as the main issue" is a device which helps to keep the thoughts directed without coercion to the point of chief interest.

Questioning. The technique of asking questions is the master key to the art of interviewing. Although the subject is given free rein at the start to tell his own story, the interviewer must eventually complete the interview by carefully selected and well-framed questions. Questions may be asked to obtain new and unexpected information which the narrative did not yield; to obtain statements either corroboratory or contradictory as a check on the original narrative; to ascertain certain names, dates, locations, etc., which were not included in the original statement; and to obtain a more complete picture of the informant's associations, biases, outlook, prejudices, motives, and interests.

Meltzer (39), in his study of social concepts, asked questions to eke out the generalizations offered with more detailed statements. If a pupil referred to *democracy* as "government by the people," then further questioning might reveal that he meant a form of government, a form of rule, a form of group activity, or whatnot.

The degree of difficulty of questions used is obviously extremely

important. Vocabulary and sentence structure must be within the grasp of the person who is to answer. One of the major causes for misinterpretation of a question is that inquirer and respondent do not have in mind the same definitions of its terms. Before we suspect a person of untruthfulness, we should first make sure that his understanding of the questions coincides with our own.

Of all questions the suggestive or leading question is most liable to produce erroneous answers. In fact, suggestion is the greatest obstacle that the interviewer must surmount in obtaining facts. It is extremely easy so to word the question, or give it an intonation by speech or gesture as to suggest the answer which the interviewer expects or would like. Well-substantiated work on the psychology of testimony shows that the leading or suggestive question decreases accuracy of response. People are open to suggestion and easily adopt the answers that the form or manner of the question makes most easy. Because children are particularly prone to yield to leading questions, this type of question should be particularly avoided with them. There are occasions, perhaps, when the suggestive question may be used. If the interviewer has reason to believe that the subject is falsifying his testimony, he may try by the skilful use of suggestion to make him contradict himself and thus expose the falsehood. This method is employed extensively in legal practice, where evasion is common.

Choice of questions. In the choice of questions lies the core of interviewing—its soundness and effectiveness or its falsity and weakness. It might seem as though it would be a comparatively easy matter to ask questions to help gain insight into a problem in conduct, but in practice to choose between relevant and irrelevant questions taxes one's utmost knowledge and skill. The information sought must be pertinent to the factors and causes that have been potent in guiding affairs into the present impasse. Knowledge is needed not only of general correlations between social phenomena but of the forces acting in the particular case.

The difficulty of selecting pertinent questions in the interview is most apparent in employment interviewing. The candidate for a position as teacher, for instance, comes in to be interviewed by a superintendent of schools. A whole range of questions could be asked concerning training, experience, knowledge of teaching

skills, religious and political affiliations, hobbies, and interests, no one of which is known to have any substantial correlation with teaching ability. A teacher may be deficient in one or more of these factors and still be a good teacher; or may measure up to each of these requirements and fail as a teacher. Knowledge of professional procedures may be counterbalanced by a lack of understanding of or sympathy with children. The field of employment psychology has developed a number of selective devices to be used in lieu of anything better which have frequently been governed by mere absurd hunches and prejudices such as antipathy to blondes, or to members of a certain race, or to those with peculiarities of speech. As a result, choices of employees have been no more successful than chance. It seems a pity to have to admit that teachers selected on the basis of attractiveness of personality have proved practically as satisfactory as though they had been chosen on what would seem to be a sounder basis, provided of course that they possess a minimum of intelligence and training.*

This discouraging admission does not mean that significant questions cannot be framed nor that the interview method is worthless for all purposes. It does mean that the choice of questions is not the obvious matter it seems. Questions must be asked after detailed study of their significance. Furthermore, the interviewer must himself have had a long and successful clinical experience, and be alert to supplement the interview by objective techniques demonstrated to be reliable and valid.

The selection of questions by systematic study is described in the chapter on questionnaires. Though this scientific study of the significance of questions seems to point toward the use of prepared question schedules and the standardization of the interview, generalization is dangerous, and the method used should be determined by the purpose of the interview. In clinical practice it is customary to draw questions from a rather lengthy schedule with the hope of uncovering significant areas where the trouble may have originated and then to follow up these clues with more pointed questioning.

An interesting discussion has appeared in the literature of social work concerning the significance and value of the non-verbal ele-

* Tiegs, E. W., *An Evaluation of Some Techniques of Teacher Selection* (Public School Publishing Company, 1928).

ments in an interview. A person expresses himself in many other ways than by words. The intonation of his voice, facial expression, flushing, bodily attitudes and bearing, and gestures are significant because these responses indicate emotional tensions, complexes, and inhibitions. The interviewer should be aware of these subtle responses, and quick to note the place where they occur in the development of the story. Often it is possible to learn more by inference from these gestures than from direct verbal expression. Queen (51), who has brought to the attention of sociologists the importance of these subtle factors of expression, remarks that no single gesture may have much significance by itself, but all may have meaning when taken together. The interviewer should also pay particular attention to their concomitants. Evasions, confessions, stuttering, shrugs of the shoulder, sharp glances, wandering of attention, sudden shifts of the topic, rubbing the hands together, and shifting the position in the chair may not be irrelevant byplay but genuine modes of response to the topic under consideration. On the other hand, of course, these responses may indicate lack of adaptation to the interviewing situation rather than to the topics being discussed.

In the training of social workers for interviewing, the suggestion has been made that students be given definite practice in observing and recording these non-verbal factors. Those with long experience in interviewing, however, dissent from this proposal. They believe that attention to these modes of response in the end defeats its purpose by withdrawing the interviewer's attention from the personal and social problem encountered, and prevents full consideration of motives, mental mechanisms, and relationships between different parts of the story. The non-verbal factors in the interview should be observed and noted, but attention to them should not usurp the place of attention to the more vital issues in the problem itself. As in any skill, attention to the process may distract one from its end or product. The process should properly be automatic, with attention free for fixation on the goal.

Climax and dénouement. The time will come in an interview when one feels that "the secret is out," the confession has been made, and the irritating sources of conflict have been uncovered. Or, on the other hand, one may have used every appeal and touched on every possible source of conflict without being able

to break down inhibition or to establish confidence. In the first case one must not be hasty in jumping at conclusions. The "secret" may be the wrong secret, or it may not be the whole secret. The successful interviewer does not depend on revelations, hunches, and coincidences. His method exemplifies the thoroughness, system, and caution which results in tentative conclusions. In some cases it may be desirable and necessary to come to a conclusion immediately. The physician must not only diagnose but prescribe. In case of doubt, however, the best results come from postponing the decision until a fresh and impartial survey can be made when all the possibilities have been considered.

Rademacher emphasizes that not too much can be expected from a single interview. One should not be at all discouraged if the interview leads to no distinct conclusion. Sampling again is a factor here. A client has been seen in only one situation during a single interview. New forces may be at work at the next sitting. It is a queer thing that inhibitions which persist with tenacity throughout a single interview require only the lapse of an interval for relaxation to occur. As soon as the subject is away, he thinks of what he might have said and is thus prepared to meet the interviewer more than half way, perhaps, at the second meeting.

Conclusion. The conclusion usually consists of mapping out a plan of action, but since we are concerned only with the diagnostic phase of the interview this will not be discussed here. One should remember, however, to terminate the interview so that relations may be resumed later if desired. Interviewer and interviewee should part as friends. "I am glad to have had this opportunity of talking things over with you" or "Let me know if I can ever be of help to you" or "Now that we are friends, come in to see me often," together with a hearty handclasp, helps to terminate the interview favorably.

Whatever the purpose of the interview, to be of most value it should probably be followed up by others. Healy (30) suggests several short interviews so as to avoid fatigue or boredom and so that the interviewer may observe the subject on several different occasions when variations in mood can be noticed. It is well known that a first interview is often unsatisfactory and that resistance broken in the initial interview may yield good results later.

Accuracy of Interviewing

General factors related to accuracy of testimony. Several decades ago there was a considerable wave of interest and research in the accuracy of observation and testimony. Summary reviews of their earlier work may be found in Whipple's reports (67, 68, 69, 70, 71, 72, 73, 74) on the "Psychology of Testimony and Report" in the *Psychological Bulletin*. Some of the pertinent conclusions reached at that time as they pertain to interviewing will be briefly sketched here.

Interrogation leads to more error both in range and accuracy of report than does simple narration. Eye-witness evidence is more reliable than hearsay evidence. Repeating a report tends to draw out memories of the original report rather than of the experience itself, so that errors in the original report are perpetuated. Any discussion of experiences with others tends to warp and distort them. The most common errors of memory are errors of omission, rather than errors of accuracy. In general there is less error on things which are said to be very certain. The time interval between an experience and its report increases the inaccuracy of the report. The accuracy of testimony can be improved by practice. Secondary qualities such as colors and numbers can be reported less accurately than primary factors. For this reason the description of an individual's appearance is usually worthless. While excitement tends to improve observation and memory up to a certain point, beyond that point it impairs them. Excitement tends to cause misunderstanding and error in giving testimony. Men are more accurate but give less extended reports than women. Children are less accurate than adults. The intelligent are more accurate than the less intelligent.

All of these statements show tendencies or correlations rather than unvarying principles. Most of them were determined before our present accurate methods of measurement and expression in terms of correlation were available, so that an elaborated repetition of the work would be desirable. Two primary sources of error hardly mentioned in this earlier work are (a) a tendency to conceal on the part of the informant and (b) errors of interpretation on the part of the interviewer.

As a general rule the subject will not say anything to incriminate himself. If he is seeking a position, he will not tell of his

past failures or discharges. The counselor should face this fact definitely and omit possibly embarrassing questions concerning previous service, for only favorable evidence will be revealed. Pupils in school can hardly be expected to tell about their own cheating, failure to study, or bad companions, nor will they admit that they are shunned by their classmates or that their parents are cruel. People dislike to report evidence that will get them into any fuss. Most persons hesitate to report speakeasies, immorality, traffic violations, and accidents for fear of getting mixed up in the matter. The workman will not report dissatisfaction with his previous job for fear it will make him less acceptable in his new undertaking.

Again it is only too often true that the person being questioned will attempt to give answers that he thinks will please or satisfy the interviewer. He studies the interviewer's biases and point of view and tries to make his answers harmonize with them so that he may earn the interviewer's esteem. It is not merely a jest that students attempt to give answers in oral examinations which they think will please the instructor rather than portray the facts as they believe them to be. Since the person being interviewed is concerned about himself and governs his story accordingly, the interviewer must make him believe that he is serving his own interest best by telling the truth.

Finally interviews are liable to error because the interviewer must interpret the subject's story. In drawing conclusions from the evidence gained in the interview, the interviewer should be extremely cautious and free from preconceptions. Perhaps the one most potent cause of error in the interview is that most interviewers start out with a handful of hypotheses uppermost in their minds and by which they are led to select evidence to fit in with their preconceived views. A case of neurosis is variously judged to be merely a set of bad habits, a condition of the cerebral hemispheres, a sex phantasy, or an infantile regression by so many different psychiatrists who sponsor these particular hypotheses. The interviewer must be extremely careful not only in what questions he asks but in what answers he rejects as trivial or unimportant. Every possible hypothesis should be considered, and rejected only when the evidence does not give it support.

Clark (21) points out that the reticent student often tantalizes the interviewer by suggesting clues which he refuses to elucidate.

Thus a frequent temptation is for the interviewer to elaborate hypotheses out of insufficient evidence, hypotheses which so soon gather a halo of plausibility that they become confused with actual evidence and are mistaken for it. Once hypotheses and evidence become confused in seeking a solution to a problem, it may happen that no distinction between them is again perceived. Every diagnosis is a hypothesis, to be sure, but in reaching some diagnoses the mind travels too far ahead of the data at hand.

Many interviewers become prejudiced on the flimsiest of excuses. The vocational counselor rejects a candidate who has a nose like that of a disliked classmate or accepts one whose voice reminds him of an old nurse. Or he gives an important interpretation to a certain hesitancy, an inability to look a person in the eye, a slight cough, or a tendency to volubility—each insignificant in itself but unconsciously made the basis for a decision.

Again the interviewer must judge the degree to which his questions are understood and comprehended by the subject and must interpret the answers in this light. Often answers are considered to be evasions merely because the subject did not fully comprehend the question.

In general an interviewer fails to give either enough credit or enough blame to those deserving it. The tendency in interviewing is to make judgments reach a rather dead level. Clark (21) found that interviews with college students made possible prediction of their grades with a correlation of .66 and .73. But the variability of the interviewer's ratings was not as great as the variability of the actual marks. Good students were recognized as good, but were not judged as bright as they proved to be. Poor students were rated as poor, but the extent of their failure was not recognized.

Statistical evidence of the reliability of interviews is almost non-existent. Interviewing has not been subjected to experimental scrutiny or statistical validation. Interviewing has resisted standardization, and many who find value in interviewing believe that to reduce it to a form of measurement is to reduce its usefulness. It is difficult to impress some enthusiasts with the unreliability of testimony. Such people, and they are the majority, prefer to rely on their own biased, cramped, inaccurate observation and judgment rather than to resort to any more deliberate, systematic, and objective methods of collecting data. The convenience of the

478 Diagnosing Personality and Conduct

interview can be no substitute for accuracy and dependability of results. Interviewing must be subjected to the same standards of reliability that other methods of collecting data are now subjected to.

The correlations found by Clark (21) between interviewers' estimates of students' ability and their actual marks (.66 and +.73) indicate that interviewing possesses some value. An interesting experiment has been reported by Barnes and Pressey on the value of oral examinations conducted in a college class. The class was split up into committees, and each of six candidates was examined in turn by the committees. This experiment showed an average correlation of .30 between the ratings of different committees. The average correlation in this instance with class marks was .47.

Industrial psychologists have demonstrated again and again the impossibility of selecting employees by interviewing, of whatever kind. The following experimental results taken from Hollingworth (31) are typical.

Fifty-seven applicants for salesmanship positions presented themselves for examination by a variety of methods.

"Twelve sales managers agreed to interview the applicants individually and to rate them for their suitability for the position in question. The interview procedure was not prescribed but was left to the dictates of the interviewer. Each judge was required finally to rate all 57 candidates according to suitability, and these ratings were then transcribed into rank from 1 to 57. The following table shows typical results."

TABLE 84
RANKS ASSIGNED APPLICANTS BY INTERVIEWERS
(from Hollingworth, 31, p. 116)

Applicant	SALES MANAGERS											
	1	2	3	4	5	6	7	8	9	10	11	12
A	33	46	6	56	26	32	12	38	23	22	22	9
B	36	50	43	17	51	47	38	20	38	55	39	9
C	53	10	6	21	16	9	20	2	57	28	1	26
D	44	25	13	48	7	8	43	11	17	12	20	9
E	54	41	33	19	28	48	8	10	56	8	19	26
F	18	13	13	8	11	15	15	31	32	18	25	9
G	33	2	13	16	28	46	19	32	55	4	16	9
H	13	40	6	24	51	49	10	52	54	29	21	53
I	2	36	6	23	11	7	23	17	6	5	6	9
J	43	11	13	11	37	40	36	46	25	15	29	1

Applicant C is given fifty-seventh place by one sales manager and first place by another. Applicant G is given fifty-fifth place by one and fourth place by another. The entire lack of consistency shows the utter impossibility of reaching agreement on personal qualifications through the interview method. Hollingworth emphasizes that these ratings were made by expert sales managers who were experienced interviewers.

Recording results. It seems to be rather generally agreed that note-taking during the interview is inadvisable. Note-taking is diverting to both the interviewer and the person interviewed. The subject's confessions are likely to be constrained if he feels that what he says is being recorded. An interviewer may be excused if he stops to record dates, names, and addresses which would be likely to escape his memory. However, the full report of the interview should be written down as soon after the interview as possible. Even if the interviewer wishes to dictate to a stenographer the report of a home visit on returning to his office, full notes should be taken immediately after the interview. Important details may be forgotten in a few minutes unless jotted down.

Many writers urge that the interview be recorded with the exact language used. So far as possible the report should be in the first person with exact quotations of the conversation. The answer to a question should be preceded by the question itself.

To be most helpful, the interviewer should attempt to include his reflections and impressions of the interview, being careful that these may be sharply distinguished in the report from the actual words spoken. The setting and background of the interview must be described. The subject's bearing, clothing, mannerisms, and approach should all be recorded. The report should reflect the accents and emphasis of the person being interviewed.

Recently social workers have been experimenting with more elaborate forms for reporting interviews. The "process interview" attempts to record not only the overt happenings and events taking place during the interview, but also the result of probing underneath the surface to describe the psychological forces at play. Elaborate forms have been devised consisting of parallel columns in which may be placed not only the events and conversations in the interview but the motives and purposes which underlie each question put by the interviewer and each answer by the subject. Such a report attempts to show in high light the mechan-

isms that carry the interview forward. This form of report is believed to be of aid in training interviewers by making them aware of their procedure and helping them to take each step in the light of the psychological forces working and the ends to be gained.

This "process interview" report may be criticized on several grounds. In the first place, whatever values it may have for training, it is a lengthy and difficult report to make and cannot be considered as practicable for clinical purposes. In the second place, from our knowledge of the difficulties of divining motives and mental mechanisms, it seems extremely doubtful that most of the interpretations can be more than guesses. The practical worker's desire for insight and interpretation must constantly be checked by the psychologist's demands for accuracy and reliability. The wish to be able to divine underlying motives in the interview should not outweigh the knowledge that such interpretations are to be considered in most cases as mere speculations.

The literature of social work also refers to the "diagnostic summary" as an integral part of the report. This summary gives the interviewer's interpretation of the interview. It emphasizes the focal points and organizes them in such a way as to present a connected analysis and picture of the situation. Such an interpretation or "diagnostic survey" ought to be included in the report of every interview. But this summary should be frankly recognized as a subjective interpretation—a hypothesis the value of which depends solely on the adequacy of the data on which it is based and the interviewer's sagacity in drawing conclusions.

REFERENCES

1. BAIN, R., "The Impersonal Confession and Social Research," *Journal of Applied Sociology*, 9: 356-361 (1925).
2. BARNES, E. J., and PRESSEY, S. L., "The Reliability and Validity of Oral Examinations," *School and Society*, 30: 719-722 (Nov. 23, 1929).
3. BINGHAM, W. V., and FREYD, M., *Procedures in Employment Psychology* (A. W. Shaw Company, 1926).
4. BINGHAM, W. V., and MOORE, B. V., *How to Interview* (Harper & Brothers, 1931).
5. BLAKE, M. B., "The Technique of the Student Interview," in Allen, F. J., *Principles and Problems of Vocational Guidance* (McGraw-Hill Book Company, 1927).

6. BLAKESLEE, L. M., "Revised Form of the Process Interview Record Form," *Journal of Delinquency*, 12: 144-187 (June, 1928).
7. BOGARDUS, E. S., *The New Social Research* (Jesse Ray Miller, 1926).
8. BOWERS, G. A., "Issues in Research Method," *Journal of Personnel Research*, 4: 155-161 (1925).
9. BRADSHAW, F. F., "The Interview: A Bibliography," *Journal of Personnel Research*, 5: 100-103 (1926-1927).
10. BRISLEY, M. S., "An Attempt to Articulate Processes," *Family*, 5: 157-161 (1924-1925).
11. BUELL, B., "Interviewers, Interviewees, and Interviewing," *Family*, 6: 86-90 (May, 1925).
12. BURGESS, E. M., "What Social Case Records Should Contain to Be Useful for Sociological Interpretation," *Social Forces*, 6: 524-532 (1925).
13. BURT, C., *The Young Delinquent* (London, University of London Press, 1925).
14. BURTT, H. E., *Principles of Employment Psychology* (Houghton Mifflin Company, 1926).
15. CADY, H. M., "On the Psychology of Testimony," *American Journal of Psychology*, 35: 110-112 (1924).
16. CAVAN, R. S., "Interviewing for Life-History Material," *American Journal of Sociology*, 35: 100-115 (July, 1929).
17. CHARTERS, W. W., and WHITLEY, I. B., *Analysis of Secretarial Duties and Traits* (The Williams and Wilkins Company, 1924).
18. CHARTERS, W. W., "Collecting of Unrecorded Specifics," *Journal of Educational Research*, 5: 280-294 (1922).
19. ———, *Curriculum Construction* (The Macmillan Company, 1923).
20. ———, "A Technique for Trait Analysis," *Journal of Educational Research*, 10: 95-100 (1924).
21. CLARK, E. B., "Value of Student Interviews," *Journal of Personnel Research*, 5: 204-207 (1926).
22. CRAWFORD, C. C., *The Technique of Research in Education* (University of South Carolina, 1928).
23. DAVIS, K. B., *Factors in the Sex Life of Twenty-two Hundred Women* (Harper & Brothers, 1929).
24. GAW, E. A., "Techniques Used in the Office of a Dean of Women," *Educational Research Bulletin*, 9: 307-310 (May 28, 1930).
25. GRAY, W. S., and MONROE, R., "Methods Employed in Making Case Studies," in *Reading Interests and Habits of Adults* (The Macmillan Company, 1929), pp. 129-140.
26. GRIFFITTS, C. H., *Fundamentals of Vocational Psychology* (The Macmillan Company, 1924).

482 Diagnosing Personality and Conduct

27. HAMILTON, G. V., *Research in Marriage* (Albert & Charles Boni, 1929).
28. HAMILTON, G. V., and MACGOWAN, K., "Marriage and Love Affairs," *Harper's Magazine*, 157: 277-287 (Aug., 1928).
29. HARRINGTON, H. F., and FRANKENBERG, T. T., *Essentials in Journalism* (Ginn and Company, 1912).
30. HEALY, W., *The Individual Delinquent* (Little, Brown and Company, 1924).
31. HOLLINGWORTH, H. L., *Vocational Psychology and Character Analysis* (D. Appleton and Company, New York, 1929).
32. KELLY, R. W., *Hiring the Worker* (The Engineering Magazine Co., 1918).
33. KIMBLE, G. E., "Self-Consciousness on the Part of the Interviewer and Its Dangers," *Social Forces*, 6: 565-567 (June, 1928).
34. LAIRD, D. A., *The Psychology of Selecting Men* (McGraw-Hill Book Company, 1925, 1927).
35. LINK, H. C., *Employment Psychology: The Application of Scientific Methods to the Selection, Training, and Rating of Employees* (The Macmillan Company, 1922).
36. LUNDBERG, G. A., *Social Research; a Study in Methods of Gathering Data* (Longmans, Green & Co., 1929).
37. LYTTON, M. C., "Art of Interviewing," *National Educational Association Proceedings*, 1929, pp. 371-376.
38. MAULSBY, W. S., *Getting the News* (Harcourt, Brace and Company, 1925).
39. MELTZER, H., *Children's Social Concepts: A Study of Their Nature and Development*, Teachers College Contributions to Education, No. 192 (1925).
40. MOORE, B. V., "The Personal Interview, an Annotated Bibliography," *Reprint and Circular Series of the Personnel Research Federation*, No. 13 (Personnel Research Federation, 1928).
41. MOWRER, E. R., *Domestic Discord: Its Analysis and Treatment* (University of Chicago Press, 1928).
42. MYRICK, H. L., "Interviews: A Study of the Methods of Analyzing and Recording Social Case Work Interviews," *Studies in the Practice of Social Work*, No. 1 (American Association of Social Workers, 1928).
43. ———, "Non-Verbal Elements in the Interview," *Social Forces*, 6: 561-564 (June, 1928).
44. ———, "Psychological Processes in Interviewing," *Family*, 7: 25-29 (March, 1926).
45. O'ROURKE, L. J., *A New Emphasis in Federal Personnel Research and Administration* (United States Civil Service Commission, 1930).

46. O'ROURKE, L. J., "Measuring Judgment and Resourcefulness as Interview Technique," *Personnel Journal*, 7: 427-440 (April, 1929).
47. PALMER, V. M., *Field Studies in Sociology: A Student's Manual* (University of Chicago Press, 1928).
48. PARK, R. E., "A Race Relation Survey," *Journal of Applied Sociology*, 8: 195-205 (1924).
49. PERSONNEL RESEARCH FEDERATION, "News Notes: The Interview," *Journal of Personnel Research*, 5: 121-122 (1926).
50. POINTER, R., *A Comparison of the Questionnaire and Personal Interview Method of Sex Investigation*, unpublished master's essay, Columbia University, 1928.
51. QUEEN, S. A., "Social Interaction in the Interview: An Experiment," *Social Forces*, 6: 545-558 (June, 1928).
52. RADEMACHER, E. S., "The Single Interview," *Mental Hygiene*, 13: 81-92 (Jan., 1929).
53. RANNELLS, M. E., "The Psychiatric Social Worker's Technique in Meeting Resistance," *Mental Hygiene*, 11: 78-123 (1927).
54. RICE, S. A., "Contagious Bias in the Interview—A Methodological Note," *American Journal of Sociology*, 35: 420-423 (Nov., 1929).
55. RICHMOND, M. E., *Social Diagnosis* (The Russell Sage Foundation, 1917).
56. ROBINSON, V. P., "Some Difficulties in Analyzing Social Interaction in the Interview," *Social Forces*, 6: 558-561 (June, 1928).
57. SALSBERY, P., "Techniques in Case Work," *Family*, 8: 153-157 (1927).
58. SCOTT, W. D., and CLOTHIER, R. C., *Personnel Management: Principles, Practices and Point of View* (A. W. Shaw Company, 1923).
59. SHEFFIELD, A. E., *The Social Case History, Its Construction and Content* (The Russell Sage Foundation, 1920).
60. SIMONS, A. M., *Personnel Relations in Industry* (Rand Book Store, 1921).
61. SNEDDEN, D. S., *Disguised Intelligence Tests, Interview Form*, Teachers College Contributions to Education, No. 291 (1927).
62. STERN, W., "Lectures on the Psychology of Testimony and on the Study of Individuality," *American Journal of Psychology*, 21: 270-282 (1910).
63. STURTEVANT, S. M., and HAYES, H., "The Use of the Interview in Advisory Work," *Teachers College Record*, 28: 551-562 (1927).
64. TAFT, J., "The Function of a Mental Hygienist in a Children's Agency," *National Conference on Social Workers Proceedings*, 1927, pp. 392-399.

484 Diagnosing Personality and Conduct

65. TEAD, O., and METCALF, H. C., *Personnel Administration: Its Principles and Practice* (McGraw-Hill Book Company, 1920).
66. WANNAMAKER, C., "The Recreation Interview in Social Case Treatments," *Family*, 6: 211-213 (Nov., 1925).
67. WHIPPLE, G. M., "A Survey of the Psychology of Testimony," *Psychological Bulletin*, 6: 153-170 (1909).
68. _____, "Psychology of Testimony," *Psychological Bulletin*, 8: 307-309 (1911).
69. _____, "Psychology of Testimony and Report," *Psychological Bulletin*, 9: 264-268 (1912).
70. _____, "Psychology of Testimony and Report," *Psychological Bulletin*, 10: 264-268 (1913).
71. _____, "Psychology of Testimony and Report," *Psychological Bulletin*, 11: 245-250 (1914).
72. _____, "Psychology of Testimony," *Psychological Bulletin*, 12: 221-224 (1915).
73. _____, "Psychology of Testimony," *Psychological Bulletin*, 14: 234-236 (1917).
74. _____, "The Obtaining of Information: Psychology of Observation and Report," *Psychological Bulletin*, 15: 217-248 (1918).
75. WHITELEY, P. L., and MACGEOCH, J. A., "The Effect of One Form of Report upon Another," *American Journal of Psychology*, 38: 280-284 (1927).
76. WOODWORTH, R. S., "Psychological Experience with the Interview," *Journal of Personnel Research*, 4: 162-165 (Aug.-Sept., 1925).

Chapter XIII

PSYCHOANALYSIS

IN psychoanalysis we have reached the frontier techniques for the diagnosis of conduct. In this region much is uncharted, much unknown. Psychoanalysis is the antithesis of standardization. The method used is free association—perfectly free, without even a list of stimulus words to react to. The physician in making his analysis must depend upon all the finesse of judgment, sagacity, and interpretative ability which he can command. Psychoanalysis is an art, not a science. In psychoanalysis one cannot give a test, score the results, and turn to a manual for its interpretation. If psychoanalysis succeeds, it is through insight derived from the careful study and wide experience of the analyst; if it fails, it is because the physician cannot measure up to the demands of interpretation and insight placed upon him.

If one becomes confused in trying to understand psychoanalysis from the expositions of it in manuals and treatises, the reason may lie partly in the fact that psychoanalysis has a threefold nature—it is first of all a *theory* of human behavior, second a technique of *diagnosis*, and third a *therapeutic measure*. These three phases of psychoanalysis are not always clearly distinguished. Most of those writing about psychoanalysis are interested in contributing to its structure as a theory of human nature, while those who practise it wish to cure mental disease. In effecting this cure they make inquiries into the intimate personal wishes and inhibitions of the patient and interpret the resulting revelations in the light of the psychoanalytic theories familiar to them. If these interpretations are accepted by the patient as being true, the cure is effected in this very act of acceptance, apparently by helping the patient to integrate reaction systems which cannot otherwise operate in harmony.

Psychoanalysis makes use of many of the techniques described in the chapter on interviewing. Certain of the principles developed in the chapter on the free association method are also utilized.

Psychoanalysis, therefore, is the most completely integrated method among all the methods considered in this book. Yet psychoanalysis as a technique is distinctly private, personal, individual.

Psychoanalytic Theory

The technique of psychoanalysis can hardly be understood without some knowledge of psychoanalytic theory. At the risk of condensing a summary of psychoanalytic theory to the point where it will become misleading, we shall undertake a brief theoretical survey before proceeding to a discussion of the technique. Although there is more than one theory, the main stream of psychoanalysis derives its conceptions from its founder, Freud. Diverging from Freud's are variations in theory sponsored by Jung and Adler, with minor variations elaborated by Rank, Ferenczi, and others.

One major difficulty in comprehending psychoanalytic theory is the vocabulary. A weighty structure of technical terms has been erected which effectively obscures its import from the uninitiated. Part of our task then is that of providing a glossary of psychoanalytic terminology.

We are asked to recognize first of all that the driving forces in the human drama are internal. Any organism has certain cellular needs and organic irritations and instabilities that constitute the driving forces to behavior. Most of these irritations and instabilities require some sort of a readjustment with the outer environment in order to effect a reduction of the irritation or a state of greater stability. Hunger and sex are the two most important illustrations of these internal driving forces in man. The driving force of sex, which is a real physiologic force, has been termed by Freud the *libido*.

This libido or driving force demands contact or readjustment with the outside world for alleviation and satisfaction. Hunger requires food in order to be satisfied, and the individual must go out and seek his food. The demands for sexual gratification require, in their unspoiled state, outer contacts for their satisfaction.

The original means for satisfying these imperative organic desires in the infant occur long before the development of language or before experiences are apprehended by means of symbols or images. In other words, they are *unconscious*. They are instinctive

if you like, or represent very rudimentary forms of learning, but they occur without any conscious apprehension on the part of the baby. Psychoanalysts make much of the fact that our earliest reactions are *unconscious*—the marvel is that some of our reactions are attended by awareness. The overwhelming number of our reactions and adjustments are unconscious, consciousness representing a more superficial froth on the deeper currents of our daily adjustments. Practically all of the tremendous amount of postural adjustment that takes place in boxing or playing a vigorous game of tennis proceeds without conscious direction.

Freud was one of the first to emphasize the fact that even the infant has sexual drives which require and eventuate in sexual adjustments. Early manifestations of the sexual interest in the infant come from the satisfactions resulting from excitation of the so-called erogenous zones, seen first in the satisfactions of suckling and later sucking. Equally potent sources of satisfaction by the stimulation of erogenous zones appear in defecation and urination. Still later in life a similar satisfaction may be aroused in connection with stimulation of the genitals.

Freud is fond of pointing out that as sexual urges develop and find their own appropriate and unique means of gratification, these expressions tend to become fixed or *fixated*. Once a person has found a satisfying mode of sexual activity, it is practised until it becomes habitual, and there is no tendency to seek alternative objects of desire or alternate modes of satisfaction. Other fundamental drives, such as hunger, are satisfied only by adjusting to a changing environment, so that they become amenable to education. But the sexual urges often derive their satisfaction within the body itself. A baby may suck its own thumb. Again the sexual desires are satisfied in the course of other activities, as when eating, disposing of bodily wastes, and the like. In other words, they are self-contained, insulated, or parasitic and hence remain more or less immune to educative processes.

Another phenomenon emphasized by the Freudians is the taboo placed by society on all sexual activity except that leading to reproduction in wedlock. As Freud (9, p. 269) says:

"One of the most important educational tasks which society must assume is the control, the restriction of the sexual instinct when it breaks forth as an impulse toward reproduction; it must be subdued to an individual will that is identical with the man-

dates of society. . . . Experience must have shown educators that the task of guiding the sexual will of the new generation can be solved only by influencing the early sexual life of the child, the period preparatory to puberty, not by awaiting the storm of puberty. With this intention almost all infantile sex activities are forbidden to the child or made distasteful to him; the ideal goal has been to render the life of the child asexual."

Consequently we find, so far as the sex life goes, two competing systems: one, the fundamental sexual drives with their fixated modes of satisfaction; the other the scruples, taboos, restrictions, and repressions established by the societal *mores* toward all matters sexual.

These competing systems, the one the responses in quest of satisfaction to the fundamental sexual urges, the other the teaching of society, lead in later life to bitter conflict and even become the root of neuroses. For we find that not only are the primary sexual satisfactions born in unconsciousness, but they are kept unconscious, are prevented from rising to consciousness, and are repressed as unsuitable for consciousness in being vulgar, lurid, and sinful. Freud has created the fiction of a *censor* in order to dramatize this inhibitory power of the conscious life best represented by those ideals of society which aim to prevent an individual from directing his attention to his sexual activities. The individual puts them out of mind, just as in the World War the power of propaganda was so strong that charitable feelings toward the enemy were not even considered, to say nothing of being definitely entertained. But the sexual life still persists in its modes of gratification which, because they are not accepted or recognized as such, appear in reactions grotesque, bizarre, and unexplainable. A neurosis has developed whose cause the individual afflicted cannot even remotely surmise.

The individual, however, must continue to live even though burdened with two reaction systems which are thus opposed to each other. The relation of an individual to objects or persons toward which he reacts sexually without the sanction of the conscious verbal organization indicates the presence of what is known as a *complex*. Complexes exhibit themselves in two characteristic ways: one by a strong charge of emotion, the other by a reticence or inhibition which causes the individual to repress the overt expression of the sexual nature of the relation. Usually complexes

are thought of as responses to words or ideas which have an emotional tinge.

Certain of these complexes have been given special names. The *œdipus complex* (so-called from the old Greek myth) consists of the tenderness of a son toward his mother, harmless enough on the surface but with far-reaching consequences. A corresponding complex called the *Electra complex* relates to the tenderness of a daughter toward the father. The *castration complex* is the belief, early acquired by a child, that in some way his genital organs are inferior or impotent, a fear that they may become so, or the fear which results when reprimanded for having committed some sexual perversion such as masturbation.

At this point in our discussion of psychoanalytic theory we ought to introduce a number of case studies. These theories of early sexual experience, fixation, repression, the complexes, and the like are so abstract and even bizarre when described that one would never dream how the psychoanalyst can see evidences of their existence in everyday affairs. But their explanations penetrate to the warp and woof of all common experience in ways that are uncanny. Many of the interpretations of everyday human experiences made in the light of psychoanalytic theory would never occur to the novice unless specifically pointed out.

The methods used in inhibiting sexual experiences, and in attempting to explain or justify sexual behavior so that it does not appear to have a sexual significance, have led to the discovery of a number of behavioral *mechanisms*. These may be mentioned briefly.

Suppression refers to the deliberate concealment or submergence of a reaction tendency which is socially unacceptable. Later this deliberate attempt to ignore or forget socially unacceptable reaction tendencies may become habitual, whereupon it is called *repression*.

Rationalization is the attempt to justify an unacceptable mode of reaction by logic, argument, or reason. One tends to explain away an inconsistency in behavior by reasons that are trivial or beside the point. The rationalizer is a person searching for props.

When one tries to explain to a person the real cause of his inconsistency, his error, his forgetfulness, or his peculiarity, the attempt is met by *resistance*. Not only is the true relationship of the experience *repressed*, but there is *resistance* toward re-

sponding to it fully. We refuse to listen to or harbor the thought that would indicate that our reactions are not really in harmony with our professions. There is a resistance which keeps two reaction systems apart.

One of the interesting ways by which sexual reactions lose their original direction and acquire new objects of allegiance is known as *transference*. This is probably nothing more than a particular instance of the laws of learning at work in the process of reintegration or conditioning. The best example of transference is the lavishing of affection on some individual who resembles in some way another whom we already love and respond to sexually. But this transference may be negative as well as positive, and dislike or disgust can be transferred as well as affection.

Regression is a term used to describe reversion to an old and successful mode of reaction in a new situation where resistance is encountered. The youth entering business where competition is keen and discouraging maintains his collegiate interest in sport as the method of outlet which has yielded him the most recognition in the past. Sometimes the regression is back to childish or even infantile forms of behavior, so we are told by the Freudians, when the difficulty to be faced by an individual is so severe that no other alternative gives the requisite satisfaction.

Two other mechanisms are known as *identification* and *projection*. In *identification* one puts himself in place of another person and reacts as he imagines the other person would act or imagines himself acting as the other person (maybe a fictional character) does act. The small boy or girl often will attempt to imitate father or mother, or will live through experiences as though he were a real robber or Indian, or she a mother or teacher. Indeed, identification in imagination is one way we gratify our wishes for which direct overt gratification is impossible.

Projection, on the other hand, is the interpretation of the acts of others by attributing to them feelings or evaluations which one possesses himself but which he either fails to recognize or resists recognizing. One may ascribe the snobbishness or unfriendliness to others which in truth characterizes better his own behavior. One may accuse others of being gossipy or critical when he experiences difficulty in inhibiting these very tendencies in himself. A pupil in school asserts that his teacher gave him his mark when in fact he determined it himself by his own achievement.

Another mechanism is *sublimation*, the draining-off of some organic desire, such as sex, into channels of activity that receive social recognition. In adolescence, particularly, when the ripening reproductive functions add their urge to the other sexual drives that have persisted through childhood, some sort of sublimation is necessary. Society puts a ban on direct gratification except through the recognized form of wedlock. In the meantime youth must seek satisfaction through sport, dress, art, religion, parties and dances, and other activities which may temporarily act as substitute satisfactions for the more direct physical gratification. Where an attempt to gratify a drive is abandoned altogether and attention is directed toward some other source of satisfaction, the mechanism is called *compensation*. The girl who cannot gain attention through good looks or attractive clothes may try to gain recognition as a good student. The boy who cannot achieve distinction in a fraternity may go in for sport or making money. In many ways one may attempt to compensate for a real or imagined inferiority by substitute activities.

Then again *defense* and *escape* mechanisms are referred to, methods of escaping from reality by withdrawing into one's own self in imagination or day-dreaming or methods of defending oneself against some inferiority by passing as a tough, hard-boiled, dangerous youngster.

The Freudian psychoanalytical theory has received severe criticism from orthodox psychologists. It is true that since as a theory it has not received scientific verification by experimental proof, it must still be considered a hypothesis. But as a theory it is extremely lusty, and there are abundant indications that it should be considered as a *promising hypothesis* even in scientific circles, worthy of being submitted to all of the tests that vigorous thought can propose. As we shall see when we discuss psychoanalytic technique, the very nature of the theory, bound up as it is with our repressions and resistances, makes it a difficult one to test by the usual experimental approaches.

The Freudian emphasis on sex as the ultimate source of practically all mental conflicts is the one feature of psychoanalysis which is most criticized by orthodox psychologists. Irrefutable evidence justifying the criticism seems to have been obtained by Rivers, whose experience with functional neuroses developed on the battle-field during the World War led him to

believe that fear, not sex, was the driving force involved in these cases.

After all, *repression* rather than sex is the corner-stone of psychoanalytic theory. Freud's contribution to our knowledge of human behavior lies in showing how the first and natural modes of expression for normal human needs and drives may become submerged beneath a veneer of custom and respectability which alone is socially acceptable. So far as his theory goes, any persistent drive that becomes repressed may lead to a neurosis. In the World War, normal fear reactions which had been stamped in by years of experience were socially unacceptable. When the soldier did not dare to exhibit fear on the battle-field with the whole country behind looking on, regressions and other mechanisms of severe neurosis developed in consequence. In everyday life such repressions without any sexual reference are going on continually. When the child who persists in asking his mother for some toy for the possession of which he has envied a playmate, is answered, "Don't mention that again," he is repressed exactly as is the child who is shamed from continuing some sexual gratification.

Probably Freud is on the whole not far wrong in claiming that in all of his analyses the clues led eventually to a sexual complex. In normal society no other drive is repressed in a wholesale way, even from thought and conversation, as is sex. Few other drives have quite the same biological persistence and endurance. No other drive is so limited in its modes of gratification. That sex is to be found the basis of the overwhelming majority of neuroses is not to be wondered at when one considers how completely society for its own protection keeps sex submerged below the thought and action current of daylight affairs. But the same mechanisms of repression and the same mechanics of substitution are to be recognized in many of the affairs of everyday life.

A rival school of psychological theory—"individual psychology"—has been developed by Alfred Adler, originally a disciple of Freud, but long since apostate. Adler was one who could not find a sexual source for every neurosis. He developed the theory of compensation for "organic" inferiority, but later broadened his theory to include all kinds of feelings of inferiority or insecurity, both real and imaginary, as a basis for nervous disorders. *A priori*, it is not so easy to envisage the source of energy for a

"feeling of inferiority" which could result in abnormal behavior, as it is for sex with its purely physiological basis. This necessity for social appreciation, however, is a drive which becomes conditioned very early in life.

"Each of us, apparently, has built up an intricate series of responses to the way in which other people flatter, ignore, or disparage our ego or personality. These responses have their roots in infancy. Our first acts are directed toward alleviating such recurrent demands as hunger, thirst, sex, drowsiness, and cold. Some person, usually the mother or father, is *always* concerned with the production of the original satisfaction, so that we learn early to respond to other people. Habits are built up of craving their presence and attention, and of dominating or being dominated by other people in order to achieve what is desired. And parents, meanwhile, are employing a variety of forms of expression such as smiling, scowling, and later flattery and scolding, as they provide or withhold these satisfactions; these forms of expression tend to become associated with the acquirement of the satisfactions. The net result is that these expressions and actions on the part of other people tend to become associated with our satisfactions and deprivations. By the time we become adults these stimulus-response situations have become so generalized that the responses which were originally very definite expressions of satisfaction and annoyance at obtaining or not obtaining what we want have melted into a vague exaltation or depression of the ego or personality." (20)

One might say, then, that Adler's theory does not go back to the beginning, and Freud claims that when traced back ultimately all neuroses have a sexual origin. We can grant that this controversy is often trivial, at least to some extent. Certainly some of the psychoanalytic literature produces explanations that are so far-fetched and bizarre as to be palpably ridiculous. Every case of neurosis presents a picture of maladjustment when traced far enough, and the clinical diagnosis should be considered adequate when the source of the maladjustment and the reason for the repression or inadequacy is disclosed. Usually it is not necessary to seek far for the source.

Technique

The techniques for conducting a psychoanalysis are many and diverse. Each practitioner has developed his own variation. Since most analysts have medical training, the subject is received as a

patient. The analyst is alert to detect the working of subtle mental mechanisms since his endeavor is to interpret and control every contact with his patient. One writer recommends that patients be dismissed by an exit which does not lead through the waiting room in which patients are received, so that there may be no contagion of impressions from those who are leaving. The welcome given to the patient is important, since the relation of the analyst to the patient is a very important part of the technique. The analyst is urged by some to shake hands and exhibit warmth and heartiness so as to establish confidence from the start. Others recommend that the analyst be pleasant but so far as possible *neutral* in his reception so that he will not be the stimulus to any transferred emotional responses in the patient.

Most analysts spend a session or two taking a case history of the patient to include dates of important events and crises, names of relatives and friends, and other data which serve as a means of introduction and general orientation. The question of fees must also be settled in these opening sessions. Patients who receive analysis gratuitously have not proven satisfactory patients, are unpunctual or irregular, lose interest readily, and tend to terminate the treatments abruptly, so that as a rule payments are recommended, however small. An analysis must extend over a long period of time—several months at least. The necessity for this should be discussed frankly with the patient at the outset and daily consultation periods should be planned at the outset to extend indefinitely into the future. To these engagements the analyst must be strictly punctual.

The routine of the session should be studiously preserved. Although the analyst must try to gain the subject's confidence, the session should be business-like and professional. The fewer the intimacies (i.e., outside of the content of the analysis) that develop between analyst and subject, the more successful the analysis. Note-taking is generally condemned, although perhaps a few random notes to help keep names or events in mind are permissible.

Freud says that he adheres firmly to the plan of requiring the patient to recline upon a sofa, while he sits behind him out of sight. Reclining encourages relaxation and facilitates the breaking-down of tensions and inhibitions. To have the analyst out of sight removes still another disturbing stimulus for the subject. It is

also important that the subject should not perceive the impressions that his revelations make on the analyst, and the easiest way to prevent this is for the analyst to keep out of sight.

The analysis is essentially an oral free-association on the part of the subject. Such a simple and trivial method would seem particularly sterile were it not for the tendency of the associations continually to veer around in the direction of significant and distressing complexes. Freud (8, Vol. II, p. 355) says of commencing the analysis, "What subject-matter the treatment begins with is on the whole immaterial, whether with the patient's life-story, with a history of the illness or with recollections of childhood; but in any case the patient must be left to talk, and the choice of subject left to him."

One fundamental rule of psychoanalysis can best be described in Freud's own words (8, Vol. II, p. 355):

"One thing more, before you begin. Your talk with me must differ in one respect from an ordinary conversation. Whereas usually you rightly try to keep the threads of your story together and to exclude all intruding associations and side-issues, so as not to wander too far from the point, here you must proceed differently. You will notice that as you relate things various ideas will occur to you which you feel inclined to put aside with certain criticisms and objections. You will be tempted to say to yourself: 'This or that has no connection here, or it is quite unimportant, or it is nonsensical, so it cannot be necessary to mention it.' Never give in to these objections, but mention it even if you feel a disinclination against it, or indeed just because of this. Later on you will perceive and learn to understand the reason for this injunction, which is really the only one that you have to follow. So say whatever goes through your mind. Act as if you were sitting at the window of a railway train and describing to some one behind you the changing views you see outside. Finally, never forget that you have promised absolute honesty, and never leave anything unsaid because for any reason it is unpleasant to say it."

With the train of free association started, the game begins—a subtle contest between the analyst, unearthing secret and hidden complexes, and the subject, more or less unconsciously trying to avoid revealing them. There are many tricks which the subject will employ, such as spending a long time in discursive associations of no particular import, talking superficially or about immediate current affairs. Another subterfuge employed by the

subject is the preparation each day of his story whereby material is carefully selected ahead of time and that which he resists disclosing is kept secret.

Through all of these incoherent ramblings the watchful analyst is able to discern focal points which indicate complexes. These may be discovered by watching for the "complex indicators" which were described in the chapter on free association. Particularly important are contradictions, gaps in memory, supplementary material inserted when repeating the description of an incident, etc. As the associations lag, the analyst may ask for the associations that are called up by one or another word, name, or phrase mentioned previously. In the later work there may be a great deal of this winnowing and probing by calling for the associations aroused by particular settings. Where these arouse emotions, pain, blocking of any kind, the probe must digress temporarily but always working to uncover the repressed material. For the main task of psychoanalysis is to unearth by free association the submerged wishes and desires which are held repressed by the respectable, society-sensitive ego.

One very important phase of analytic technique is the analysis of the subject's dreams. In psychoanalytic theory dreams are held to be the expression of the basic but repressed desires which are distorted by various mechanisms of *symbolism*. A debated and criticized aspect of Freudian psychoanalysis is the codification of dream symbolism as though there is or tends to be a definite and necessary correspondence between a particular dream symbol or dream content, wherein it occurs, and a particular diagnostic significance or psychoanalytic connotation. The interpretation of dreams is especially complex because of the variety of mechanisms by which dream material is distorted. Freud says: "In general it is doubtful in the interpretation of every element of the dream whether it (a) is to be regarded in a negative or positive sense, (b) is to be interpreted historically, (c) is symbolic, (d) its valuation is to be based upon the sound of its verbal expression."

A psychoanalytic diagnosis must depend entirely on the sage interpretations which the analyst can make in the light of psychoanalytical theory. The analyst is always sensitive to evidences of the existence of complexes, he watches for indirect expressions of sexual gratification, the objects or persons to which it is attached,

and the degree to which there is satisfaction or unsatisfied longings and desires. The analyst watches particularly for the relationship of the subject to familiar persons in the life story as it is unfolded, to discover possible hidden sexual relationships which would never be recognized and acknowledged from everyday surface events even by the subject himself. The analyst is quick to explain otherwise colorless admissions or denials by one or another of the common "Freudian" mechanisms, thus tending to give them a hidden, secret, sexual significance.

Other features of the analysis—its length, decision as to the proper time to disclose interpretations to the patient, or the important transference to the physician of emotions of love or hate, cannot be discussed here, as they belong primarily to psychoanalysis as a therapy and not as a diagnostic technique. In fact much of the cautiousness with which an analysis proceeds is based on therapeutic considerations rather than on a regard for what is essential to the diagnosis. Most psychoanalysts maintain that they are able to divine the essential features of the case long before it is possible to reveal them to the patient.

Indeed Freud practically admits that if an understanding of the case by the physician is all that is wanted, psychoanalysis is an awkward method of attack. He says (8, Vol. II, p. 362):

"In the early days of analytic technique it is true that we regarded the matter intellectually and set a high value on the patient's knowledge of that which had been forgotten, so that we hardly made a distinction between our knowledge and his in these matters. We accounted it specially fortunate if it were possible to obtain information of the forgotten traumas of childhood from external sources, from parents or nurses, for instance, or from the seducer himself, as occurred occasionally; and we hastened to convey the information and proofs of its correctness to the patient, in the certain expectation of bringing the neurosis and the treatment to a rapid end by this means. It was a bitter disappointment when the expected success was not forthcoming."

In other words, Freud admits that if the facts are what is wanted, one need not use the cumbersome psychoanalytic method but may search and gather them wherever most available. In practical application in diagnosing a child's problems one need not question the child exclusively, but should gather evidence from teachers, parents, or friends, as most convenient.

Evaluation of Psychoanalytic Technique

Psychoanalysis as a diagnostic technique is inaccessible to validation by the standard tests which are employed in determining the reliability and validity of evidence. Psychoanalysis, since it can never be other than the intimate confession of subject to analyst, is immune to the usual scientific checks which may be used. In consequence of this, our evaluations of it must be deductive, we must fall back upon our knowledge of the worth of related techniques. To most of the criticisms leveled against it psychoanalysis has prepared a peculiar kind of answer which is invulnerable to ordinary psychological logic.

Psychoanalytic diagnosis may be criticized in the first place by questioning the accuracy or objective reference of the patient's free associations. Our knowledge of the psychology of testimony would lead us to expect that here as elsewhere there would be serious gaps or distortions of memory and observation. Where there is emotion or a complex, the interpretation of events is particularly apt to be in error. Although it is well known that memories lose their accuracy with the lapse of time, psychoanalysis makes much use of memories of childhood occurrences, with little regard for the probability of distortions. The psychoanalyst admits there are errors and lapses in observation, memory, and interpretation, but maintains that he recognizes them when they occur and properly attributes them to the mechanism which occasioned them. In fact, practically nothing that the subject discloses is taken at its face value—all is interpreted in the light of psychoanalytic theory. Freud, drawing on the results of the free association experiment, concludes that trains of free association follow certain laws and are governed by certain forces. They may coincide with reality only rarely and in dreams may be fairly fantastic, and yet they are constantly infringing on reality. Since the psychoanalyst readily admits that testimony is fact mixed with fiction, the sting of this first criticism is removed.

While considering the accuracy of the testimony one should discuss another criticism that has been raised against psychoanalytic diagnosis. To what extent does knowledge of psychoanalytic theory color and distort the subject's free associations? If a subject knows that psychoanalysis always looks for some ungratified sexual desire, and if he voluntarily undergoes analysis, does he

immediately play up to what he thinks is expected of him? This is a serious question. Few people undergo psychoanalytic treatment without having some idea of what psychoanalysis is. Does the psychoanalyst ever know when he is being humbugged by having his subject say the things which he thinks he is expected to say?

If so much depends on the interpretation which the analyst makes of the free associations of his patient, does not our knowledge of the psychology of judgment give us good ground for entirely discrediting a psychoanalytic diagnosis? If it is so difficult to rate the character of pupils, employees, and friends even after close observation, is it possible that the analyst can read the truth from the random mental wanderings of his subject by his own sheer interpretative abilities?

Psychoanalysis has been severely arraigned for its methods of arriving at conclusions which would never for an instant be admitted by science or in a court of law. Tannenbaum in a clever article gives a clear exposition of how Freud in interpreting one of the cases he describes, jumps at conclusions prematurely and with insufficient evidence. He draws one conclusion where, by his own theories, other conclusions could fairly be drawn. Tannenbaum (21, p. 62) concludes his "exposure" by stating, "It is the wilfulness, arbitrariness, and capriciousness of the Freudian technique and interpretations appearing in every instance of an 'analysis' submitted to critical tests that prove orthodox psychoanalysis to be nothing more than a cleverly conceived therapeutic system which has no more scientific validity than Christian Science."

If Freud with his genius and plausibility and ingenuousness is so vulnerable to attack, what of the less capable amateur "analysts" who try to employ his methods? Here unsubstantiated assertions are crassly in evidence. Take for instance the following statements found in Zachry's *Personality Adjustments of School Children* (27, p. 55): "It has already been suggested that Ned's scowling and tantrums are his defense against his feelings of insecurity and inferiority." Or (27, p. 115), "Esther made up for difficulties, or made an attempt to compensate for them, by substituting school success for everything that she lacked. Her failure to compensate was due, in part, to the fact that she had no intensive interest in her school work. It may be that because her

thwarted desires and repressions are concerned with her love needs, they cannot be compensated on a purely ego level, the level of her school work." Or consider such statements as are found in Rogers' (18) *Measuring Personality Adjustments in Children Nine to Thirteen*, "Long contact with a psychiatrist revealed the fact that these fears of being attacked seemed to take their root in a very early experience when he was attacked and severely bitten by a dog. . . . It seems possible that the early conditioning fear of the dog colored much of Walter's life." Or in another case: "The stealing and enuresis were more or less deliberate weapons against the aunt, who was both unsympathetic and unwise." The fashion set by Freud of interpreting all unusual or undesirable behavior in terms of adjustment mechanisms may finally result in pure speculation on the part of his followers. Even though possessing a certain plausibility, such statements as the foregoing cannot be considered as inevitable conclusions from the evidence presented.

Freud is apparently insensitive to these attacks, although he is aware of them. He possesses a boundless confidence in the accuracy of his interpretations. For instance, with regard to dreams he says (9, p. 195), "What might, for example, impress you as arbitrariness in the interpretation of symbols, is compensated for by the fact that as a rule the connection of the dream thoughts among themselves, the connection of the dream with the life of the dreamer, and the whole psychic situation in which the dream occurs, chooses just one of the possible interpretations advanced and rejects the others as useless for its purposes." Freud never forces an interpretation. The train of free associations advanced must continue until some one interpretation becomes irresistible. That is, psychoanalytic diagnosis depends for its validation on the degree of consistency and harmony which its interpretation contains rather than its obedience to the laws of inference. It is undeniable that psychoanalytic interpretations have a sort of enticing plausibility to even the superficial observer.

Psychoanalysis as a diagnostic technique should be viewed with an attitude of *favorable skepticism*. It cannot be scientifically validated and can be severely criticized when studied in the light of the psychology of testimony, judgment, and inference. On the other hand, its keen application of psychological mechanisms, coupled with a dynamic psychological theory of considerable vigor,

gives it a somewhat impressive plausibility. Psychoanalysis should be considered a *promising hypothesis*. It offers unique challenge to the psychologist of scientific bent. Perhaps the severest arraignment of the psychoanalytic method concerns its unconscionable demand for unlimited time. Its uneconomical procedure constitutes a challenge to the psychological engineer who, aware of the significance of the mechanisms it uncovers, can devise some means of breaking *resistance* more easily by methods that can be standardized and which can be employed with groups. Such devices are certain to arouse opposition at first because they probe into things ordinarily considered private and personal, and because early attempts are almost certain to be cumbersome. But this survey of the whole range of techniques for the diagnosis of conduct points to psychoanalysis as the unknown promised land.

REFERENCES

1. ADLER, A., *The Neurotic Constitution* (Moffat, Yard & Company, 1916).
2. ———, *Study of Organ Inferiority and Its Psychical Compensation* (The Nervous and Mental Disease Publishing Company, 1917).
3. ———, *Understanding Human Nature* (Greenberg, Publisher, Inc., 1927).
4. BJERRE, P., *The History and Practice of Psychoanalysis* (Richard G. Badger, The Gorham Press, 1916).
5. BONSFIELD, P., *The Elements of Practical Psychoanalysis* (London, Kegan Paul, Trench, Trubner, & Company, 1920).
6. COSTER, G., *Psychoanalysis for Normal People* (London, Oxford University Press, 1926).
7. FORSYTH, D., *The Technique of Psycho-analysis* (London, Kegan Paul, Trench, Trubner, & Company, 1922).
8. FREUD, S., *Collected Papers*, Volumes I, II, III, and IV (The International Psycho-analytical Press, 1924).
9. ———, *A General Introduction to Psychoanalysis* (Boni and Liveright, 1920; now published by Horace Liveright).
10. GLOVER, E., "Lectures on Technique in Psychoanalysis. I. The Analytic Situation. II. The Opening Phase," *International Journal of Psychoanalysis*, 8: 311-338, 486-520 (1927); 9: 7-46, 181-218 (1928).
11. GREEN, G. H., *Psychanalysis in the Classroom* (London, University of London Press, 1921).
12. HAMILTON, G. V., "Comparative Psychology and Psycho-

- pathology," *American Journal of Psychology*, 37: 200-211 (1927).
13. HEALY, W., BRONNER, A. F., and BOWERS, A. M., *The Structure and Meaning of Psychoanalysis* (Alfred A. Knopf, 1930), 482 pp.
 14. JELLIFFE, S. E., *The Technique of Psychoanalysis*, Nervous and Mental Disease Monograph Series, No. 26 (1918).
 15. LOW, B., *Psychoanalysis* (London, George Allen & Unwin, 1920).
 16. PFISTER, O., *The Psychoanalytic Method* (Moffat, Yard & Company, 1919).
 17. RIVERS, W. H. R., *Instinct and the Unconscious* (Cambridge University Press, 1920).
 18. ROGERS, C., *Measuring Personality Adjustments in Children Nine to Thirteen*, Teachers College Contributions to Education, No. 358 (1931).
 19. ROSANOFF, A. J., *Manual of Psychology*, Sixth Edition (John Wiley and Sons, Inc., 1927; First Edition, 1905), Part III, Ch. V.
 20. SYMONDS, P. M., "An Analysis of Tact," *Journal of Educational Research*, 21: 241-254 (April, 1930).
 21. TANNENBAUM, S. A., "Interpreting a Compulsion Neurosis," *Journal of Abnormal and Social Psychology*, 20: 48-62 (1925).
 22. TRIDON, A., *Psychoanalysis: Its History, Theory and Practice* (The Viking Press, 1919).
 23. VAN TESLAAR, J., *Outline of Psychoanalysis* (Boni & Live-right, 1924).
 24. WECHSLER, I. S., *The Neuroses* (W. B. Samuels Company, 1929).
 25. WOHLGEMUTH, A., *A Critical Examination of Psychoanalysis* (London, George Allen & Unwin, 1923).
 26. YELLOWLEES, H. A., *A Manual of Psychotherapy for Practitioners and Students* (The Macmillan Company, 1923).
 27. ZACHRY, C. B., *Personality Adjustments of School Children* (Charles Scribner's Sons, 1929).

Chapter XIV

EXTERNAL SIGNS OF CONDUCT

THERE is a thriving band of quacks and charlatans who frequently go under the name of "applied psychologists," but who really are pseudo-psychologists, traveling about from city to city, selling their wares in lecture halls, churches, lodges, and industrial plants, playing on the credulity of the gullible, working on the enthusiasm inspired by hope, and claiming to teach how to "read" character from its physical signs. They depend on the tendency of the uneducated to generalize from specific cases. They rely on the compliance that is usually accorded one who speaks fair, pompous words full of authority.

On the other hand there is the group of self-elected St. Georges who champion the cause of scientific psychology against pseudo-psychology, laying about with lusty blows to repel the menaces of superstition and ignorance. These defenders of the faith, alas, do not write with the calm dispassionateness of scientists, for they are so outraged by the easily won popularity of the pseudo-psychologists as to become vitriolic in their denunciations; in fact they are as intolerant in their vituperations as those whom they assail are plausible in their inferences.

Fortunately there has also been a small group of level-headed thinkers who have been willing to experiment and determine the facts, who approach the problem without bias and are willing to see relationships when they have been proven to exist, but who also demand that the claims be given fair, impartial trial on a sufficient number of cases.

The interests of those who attempt to infer or predict conduct from external signs may be classified into the following groups:

1. Events occurring outside the body.
 - a. Astrology.
2. Characteristics of the body without even a remote relationship to conduct.
 - a. Palmistry.
 - b. Phrenology.

3. Physique.
 - a. Height.
 - b. The morphologic index.
 - c. Kretschmer's types.
4. Characteristics and stigmata.
 - a. Lombroso's stigmata.
 - b. Estimating character from physiognomy.
 - i. Facial profile.
 - c. Estimating character from photographs.
 - d. Blonde-brunette types.
 - e. Shape of hand.
5. Handwriting.

Even this list omits some of the most fantastic of character systems, such as numerology, iridology, and fortune-telling from playing-cards or tea-leaves.

Events occurring outside the body that have no possible relation to one's destiny may be dismissed immediately as unworthy of our attention. It seems strange that in this enlightened age masses of people not only in India but also in the western hemisphere, America included, still place their faith in astrology. At one time astrology so influenced the "scientists" of the day that such a word as *lunatic* even became a part of the language. But to-day we recognize that the planets and signs of the zodiac have not even the remotest influence in guiding human affairs.

With almost equal abruptness we may dismiss the two pseudo-scientific systems, *palmistry* and *phrenology*. There is a certain type of mind which loves to make much of the trivial and to find in it occult portents. The significance of the lines of the hand, which in sober fact correspond to natural divisions between muscle groups, has been magnified until some have found in them a complete system of character analysis. The famous anatomist Alexander MacAlister * says of this,

"That these mechanical arrangements have any psychic, occult, or predictive meaning is a fantastic imagination, which seems to have a peculiar attraction for certain types of mind, and as there can be no fundamental hypothesis of correlation, its discussion does not lie within the province of reason."

Phrenology is a system which undertakes to diagnose character by means of swellings or protuberances on different parts of the skull. This system, first formulated by F. J. Gall about 1800, was

* *Encyclopedia Britannica*, 14th Edition, Vol. XVII, article "Palmistry."

due to a false conception of the function of the brain. It was assumed in a crude form at that time that the different sections of the brain are the seats of different mental "faculties" or "powers." Consequently there was supposed to be a correspondence between the development of certain sections of the brain (which could be noted by studying the external contour of the skull) and certain faculties or personality characteristics of a person. Gall went to work empirically (on a few cases) to locate the seats of these faculties. Charts and diagrams have been developed mapping the location of such characteristics as love of approbation, cautiousness, combativeness, destructiveness, the perception of time, tune, and number, and the like.

So far as can be shown, this system does not deserve the least amount of serious consideration. In the first place, it has long since been demonstrated that the brain does not operate functionally in any way that corresponds to the system of phrenology.

Indeed, the recent work of Lashley indicates how far we have left behind this view that there is a necessary localization of any *complex* function (such as the phrenologists always talk about) in any part of the brain, one part assuming the work and function of another part if necessary. The system of phrenology is also false in that the "mental faculties" on which the system is based have long since been shown not to exist. Human behavior is a dynamic system of response to stimulation and is not an expression of certain faculties. MacAlister * likewise disposes of phrenology as follows: **

"Psychology, physiology, and experience alike contribute to discredit the practical working of the system and to show how worthless the so-called diagnoses of character really are. Its application by those who are its votaries is seldom worse than amusing, but it is capable of doing positive social harm, as in its proposed application to the discrimination or selection of servants and other subordinate officials."

Physique as Diagnostic of Conduct

The systems which propose to diagnose personality by a superficial survey of a man's physique are many. It is only natural that

* *Encyclopædia Britannica*, 14th edition, Vol. XVII, article "Phrenology."

** See also Paterson, D. G., "Personality and Physique," in *The Measurement of Man* (University of Minnesota Press, 1930).

there should be postulated a relationship between the superficial, outward characteristics of size and shape of the body and the expression of this body in conduct. These claims must be given some serious consideration.

One of the most natural of such claims is that tall men have certain possibilities for aggressiveness or leadership not accorded to small men. It has been claimed that tall men make more successful salesmen, that tall men are chosen to be leaders by the members of a group, and that height is an advantage wherever there is a call for dominance over others. It is a common observation that school superintendents when together in a group appear to be above average in height. Yet to cite one example on the other side, Napoleon achieved prominence as a leader, although only a very short man.

Several investigators have studied this asserted relation and provide us with a check on it. Gowin (27) obtained the heights and weights of 6,037 leading men such as governors of States, United States senators, mayors of leading cities, bishops, railroad presidents, etc. The average height of this group was 71.4 inches and the average weight was 181.1 pounds. These figures may be compared with 221,819 applicants for life insurance, whose average height was 68.5 inches and average weight 166 pounds. In Gowin's study, those who were in more responsible positions exceeded those in less responsible positions in both height and weight. Kitson, on the other hand, in studying the relationship between success in salesmanship and height, found that the taller salesmen were no more successful than the shorter. Kitson says (37, p. 92),

"Thus we see that the salesmen investigated are not appreciably taller than men in general. While the extremely tall salesman may in the long run possibly excel the extremely short salesman, he will not excel the salesman of average height. Indeed, the figures show that he will not do so well as the man of medium height. Accordingly, we should not be justified in looking to height as a physical criterion of success in salesmanship."

Bellingrath (3), in a study of school leaders as determined by office-holding in extra-curricular activities in high school, finds that the average height of seventy-four leaders was 67.7 inches against 65.9 inches for those in the same classes who had not held office.

The leaders had an average weight of 140 pounds, while the non-leaders had an average weight of 129 pounds. These differences were more pronounced in the girls than the boys. In the case of the girls, biserial r 's of .44 in height and .42 in weight were found.

We must conclude that height and weight are positively correlated with leadership and the accompanying characteristics of leadership. But the overlapping is considerable and one would not be safe in picking candidates for positions of responsibility on the basis of height and weight alone.

Two lines of investigation into the significance of bodily build have been followed. One of these was pursued vigorously by Sancti Naccarati (48, 49, 50) under the caption of the "morphologic index." The other centers in and is largely due to the stimulating suggestions of the German psychiatrist Kretschmer (42) concerning the relations of physique and character.

Naccarati entered upon his problem with a background of knowledge of the work of the Italian anthropologists, particularly Viola. Naccarati recognizes two extreme morphologic or structural types. (Morphology is a biological science dealing with the form and structure of plants and animals.) *Microsplanchnics* are individuals with small trunks as compared with the development of their limbs. These are the persons who strike one as being long and thin. The *macrosplanchnics*, on the other hand, are those with large trunks as compared with the development of their limbs. They impress one as being stout or fat. Naturally it is recognized that these are merely the extremes of a continuous gradation and that the largest group of men are *normosplanchnics*. Naccarati believes that these two extreme types represent fundamental differences in human temperament. The macrosplanchnics represent those in whom the nutritional and vegetative systems, in which energy is stored up, have had the most extensive development. The microsplanchnics represent those in whom the systems of locomotion and manipulation by which energy is transformed and utilized are in the ascendancy. He believes that this latter type is correlated with a higher mental organization. Physiology identifies the microsplanchnic type with hyperthyroidism.

Naccarati developed a "morphologic index" for accurately measuring these differences in body types. The somewhat complex formula for computing the morphologic index is as follows:

$$\text{Morphologic index (M. I.)} = \frac{\text{length of two limbs}}{\text{volume of trunk}}$$

More in detail the computation is carried out as follows (68, p. 448):

1. Sternum length.
2. Xipho-epigastric line.
3. Pubo-epigastric line.
4. Transverse thoracic diameter (width of upper chest).
5. Anterior-posterior thoracic diameter (depth of upper chest).
6. Transverse epigastric diameter (width of upper chest).
7. Anterior-posterior epigastric diameter (depth of lower chest).
8. Transverse pubic diameter (width of waist).
9. Length of lower extremity.
10. Length of upper extremity.
11. Height.
12. Weight.

- a. $1 \times 4 \times 5 = \text{thoracic index (volume)}$
- b. $2 \times 6 \times 7 = \text{index of upper abdomen (volume)}$
- c. $3 \times 7 \times 8 = \text{index of lower abdomen (volume)}$
- d. $a + b + c = \text{trunk-value (volume)}$
- e. $9 + 10 = \text{limbs value (combined length)}$
- f. $e \div d = \text{M. I. (morphologic index)}$

The larger the M. I., the greater the tendency toward the micro-splanchnic type.

With all of this set-up and in spite of the refinements of these measurements, the results have been disappointing. Naccarati himself found a correlation of .35, with seventy-five Columbia College students, between the M. I. and the Thorndike "Intelligence Test for High School Graduates." Sheldon (68), however, repeating the experiment in Chicago, found a correlation of only .14 between the M. I. and the American Council Intelligence Test.

Heidbreder (29), in investigations on college students, finds a correlation of only $.03 \pm .03$ between the height-weight ratio and score on the "Minnesota College Ability Tests" for 500 men, and $.04 \pm .03$ for 500 women.

Naccarati and Garrett (49) found practically no relationship—and were later substantiated in this by Garrett and Kellogg (25)—between the M. I. and the Woodworth personal data sheets, the Pressey X-O tests, or ratings. In short, they were not able to demonstrate that there is any connection between the M. I. and temperamental qualities.

In this connection work done by Paterson under the direction of Heibredner and Paterson at Minnesota indicates the probability of a slight relationship between physique and temperament. Using the Heibredner "Inferiority Attitude Self-Rating Scale," a correlation of $+.03 \pm .03$ was found for 673 college freshmen between the height-weight ratio and the inferiority attitude score. Using 531 freshmen, a correlation of $+.11 \pm .03$ was obtained, the correlation of inferiority-superiority with height being $+.10 \pm .03$ and with weight $-.10 \pm .03$. As Paterson (55) sums up, "Apparently there is a slight tendency for those who are tall and thin to rate themselves higher on the scale," i.e., toward feelings of inferiority.

Sheldon (69), however, found that the M. I. correlates $-.22$ with ratings on sociability and $-.14$ with ratings on leadership. Sheldon concludes from his experimental survey of the field that "the factor of general size, or bigness, seems to be related positively to sociability, leadership, and aggressiveness," and his evidence shows the possibility that a slight relationship may exist.

Kretschmer (42) worked on practically the same problem from the psychiatric point of view, but he developed a new terminology which requires that his work be treated separately. Kretschmer recognizes four types of individual, called respectively the *asthenic*, the *athletic*, the *pyknic*, and the *dysplastic* forms. These forms are to be distinguished from each other by their bodily build, by the muscular development and placement of fat on the body, by the shape of the face and hands, and by the character of the skin and hair. These four types are succinctly characterized by Mohr and Gundlach as follows (47, pp. 118, 119):

"The *asthenic* individual, said to be of the schizophrenic temperament, is one who is of average height but is relatively tall for his weight. He is thin, with a long, narrow, shallow chest. His shoulders are relatively broad contrasted with the diameter of his chest. His muscles are thin and poorly developed, his skeletal structure is slight. The skin is thin and loosely attached to the underlying tissues. The face is characteristically long and narrow, with a prominent nose and clear-cut features. The facial angle is sharp, and the mid-face is relatively long.

"The *athletic* physique is similar to the *asthenic* in general bodily proportions but all of the structures are thicker, firmer and of more robust development. The shoulders are heavy, the chest is broad and of medium depth. The skeleton is heavily built. The muscles are thick, of good tone and are well contoured. The skin

is thick and closely adherent. The face is relatively long and narrow with proportions similar to the asthenic but with thick though well-defined features. The facial angle is less marked than in the asthenic and the lower jaw more heavily developed. The athletic build is considered a variant of the asthenic.

"The *pyknic* habitus is described as one in which there is an increase in the volume of all of the body cavities. The head is large, the chest is voluminous and exceptionally broad and deep. Although the shoulders are of moderate width they appear narrow contrasted with the broad chest. The abdomen is full. The skeletal structure is slight when compared with the general bulk of the individual and the extremities are relatively small and slender. The hands are small and delicate. There is a generous adiposity and the skin is thick and firm. The face is round and the midface is short. The complexion is ruddy.

"The *dysplastic* type includes many deviants from the normal. In this group are those physical forms evidencing disturbance of the various ductless glands. The elongated form of the eunuchoid, hypoplastic forms and those in which there have been localized developmental disturbances, are included."

Kretschmer believes that there is a marked relationship between these bodily types and temperamental qualities. In his volume, *Physique and Character*, he draws with the enthusiasm of the clinical worker a clear-cut picture implying a degree of relationship higher than it really is.

His studies with the insane have led him to the view that there are two general types of functional insanity. These two types also exist in minor degree in normal persons, as shown by what he calls the *cyclothymic* and *schizophrenic* temperaments.

The *cycloids* (having the cyclothymic temperament) include those who wear their emotions on their sleeves. They are extremely expressive and tend to fluctuate in mood from one of joyful excitation to extreme depression. The cycloid as we know him in daily contacts is the sociable, good-natured, friendly, genial, hail-fellow-well-met sort of an individual. He takes life as it comes, gets on well with people, and is ready to understand a joke. On some occasions he is cheerful, jolly, hasty, while at other times he is quiet, calm, soft-hearted, and easily depressed. The fat boy in *Pickwick Papers* and the irrepressible Mr. Micawber in *David Copperfield* are examples of the type. Naturally one finds them with all degrees of variation.

The *schizoid* (having the schizophrenic temperament), on the

other hand, is an unsociable, quiet, reserved, serious, eccentric individual. His emotional life is hidden within. There is a barrier between him and society. He is the man that no one knows or understands. Only at rare intervals does the inner man break through, revealing the depth of feeling hidden within him. On occasion he will be timid, shy, sensitive, showing fine feelings, nervous, excitable, fond of nature and books. Again he will be found to be stolid, pliable, kindly, honest, indifferent, dull-witted, silent.

Kretschmer believes there is a marked relationship between these two types and differences of physique. The *asthenic* and *athletic* types of bodily build go with the *schizoid* temperament. The tall, thin man tends to be the reserved, sensitive, unemotional individual. On the other hand, the *pyknic* types go with *cycloid* temperaments. The fat, well-developed person is also the one who is genial, friendly, and easy to meet. Kretschmer gives the following table to show the extent of the relationship:

TABLE 85
RELATION BETWEEN TEMPERAMENTAL TYPES AND PHYSIQUE
(from Kretschmer, 42, p. 35)

<i>Physique</i>	<i>Cycloid</i>	<i>Schizoid</i>
Asthenic	4	81
Athletic	3	31
Asthenic-athletic mixed	2	11
Pyknic	58	2
Pyknic mixture	14	3
Dysplastic	34
Deformed and uncataloguable forms.....	4	13
	<hr/> 85	<hr/> 175

Kretschmer's claims have received confirmation in Europe and America.

Wertheimer and Hesketh (79), working in Baltimore hospitals, present confirmation of Kretschmer's hypothesis. Eleven cases of manic-depressive insanity had an index of 233 (a low score indicating the stout build), whereas twenty-three schizophrenics had an average of 281 (indicating a tendency toward longness and thinness), but there was much overlapping. From the table of their results it is evident that there is a close relation between the tendency toward the manic-depressive psychosis and the

512 Diagnosing Personality and Conduct

pyknic body build, but that the schizophrenic cases although tending toward asthenic or athletic types are not so closely confined to one type.

TABLE 86

SHOWING THE PERCENTAGE DISTRIBUTION OF MANIC-DEPRESSIVES AND SCHIZOPHRENICS ACCORDING TO KRETSCHEMER'S BODY TYPES

(after Wertheimer and Hesketh)

Body type	CLEAR MANIC-DEPRESSIVES		CLEAR SCHIZOPHRENICS	
	N = 11. All ages	N = 6 Age 20 +	N = 23 All ages	N = 12 Age 20 +
Pyknic	45.5	66.6	4.3	8.3
Pyknoid	36.4	33.3	13.0	25.0
Athletic	9.0	0	26.1	16.7
Asthenic-athletic mixed ..	0	0	34.8	25.0
Asthenic	0	0	17.4	16.7
Nuclear	9.0	0	4.3	8.3
	<hr/> 99.9	<hr/> 99.9	<hr/> 99.9	<hr/> 99.9

Mohr and Gundlach's (47) study concerned primarily the *performance* of criminals who had been classified according to Kretschmer's types, and they obtained social data indicative of the temperament and personality of the groups. Nineteen asthenic criminals as compared with forty-four pyknic criminals showed a larger percentage of crimes against property and a small percentage of crimes against the person. The asthenic group had a much lower record of previous imprisonment. A smaller percentage of the asthenic group were married and a smaller percentage belonged to fraternal organizations. These groups differed in race, age, and social status, which may impair to some extent the clear-cut character of the findings. All of these findings, however, are in harmony with Kretschmer's theories regarding the temperamental differences between the two groups.

Farr (20), working in the Pennsylvania Hospital, reports studies of the relationships between anthropometric measurements and types of psychosis. Very careful measurements were made of a group of sixty patients, including eleven characterized as schizoid and sixteen as affective. Farr concludes from his studies (20, pp. 235, 236):

"The figures suggest a rather definite association of seclusive and schizoid personalities with the slender, relatively elongated types—often with dysplastic features—and of the affective per-

sonalities with intermediate or definitely thick-set physiques. This is entirely in agreement with the observation of others, but the outstanding exceptions and the questionable correlations are so numerous that anthropometry must be looked upon rather as interesting and suggestive than as diagnostic."

One of the difficulties with Kretschmer's system is that he has given only a descriptive characterization of the types; he has not defined them rigorously in quantitative terms. With this loose kind of classification instead of measurement to go by, investigators have formulated widely different criteria of the various types or have placed different amounts of emphasis on the criteria (physique, face, hair, skin, etc.), so that there is wide variability in the incidence of the types among insane patients studied by different investigators. Some have found it difficult to distinguish between the athletic and asthenic groups. Kretschmer in recent writings would combine these two groups into a single *leptosome* type, but Mohr and Gundlach do not believe this is justifiable.

Although Kretschmer used measurements to describe his types after he had selected them, he apparently is not willing to advocate making the original selection on the basis of measurements. He says (42, p. 13):

"It is superfluous [to procure statistical values for every single measurement of the body] because even within a single people we can give only approximate figures, since, when dealing with body-forms, we have not to do with clearly marked unities but with ill-defined types, where, in the case of certain border-line cases, it depends on the investigator himself whether he includes them on the list of the one single type or not." [And in another place] "The important idea about a type is that it possesses a firm center but not hard and fast boundaries. Types as a rule can only be determined intrinsically; we cannot mark their boundaries. By 'type' we mean a nucleus of more distinct and among themselves quite firm formations which have been deliberately lifted out from a sea of progressive transitions. This holds good for a racial type as well as a personality or a clinical reaction type."

The breakdown of Kretschmer's classification system is due to his attempt to approach the problem through distinctions of types. Types even when they do not violate logic are not reconcilable with statistics and the fundamental principles of measurement. The truth of the matter is that all of the qualities which go to

characterize a type can be measured on a continuous scale of more or less. It is probable that all of them would form distributions approximating the normal probability curve, with few at the extremes and many near the median. Each of these qualities (body dimensions, facial contour, skin texture, hair) has more or less correlation with temperamental differences. But they are also more or less imperfectly correlated among themselves, so that it is seldom that an individual illustrates a pure type. The problem of using physique in the diagnosis of conduct thus becomes a statistical problem of correlation. The task of isolating the significant variables and combining them with the best weighting for the diagnosis of temperamental differences is a problem yet to be worked out.

Enough has been done in this field to show that there are certain relationships between physique on the one hand and personality on the other that have diagnostic value. The clinical experience of such psychiatrists as Kretschmer has uncovered promising leads. But the scientific work of determining the exact relationships and making use of these relationships to fashion a measure of diagnostic significance yet remains to be done. Even when it is done, it cannot be used as a complete measure of temperamental differences, for there will still remain such factors as age, environmental stress and strain, etc., to accentuate or reduce the clear-cut temperamental manifestations.

Bodily Characteristics and Stigmata as Diagnostic of Conduct

Our discussion cannot come to an end when we have taken up general physique only, for every conceivable characteristic of the human body has at one time or another been believed to be indicative of tendencies in conduct. Many of these imaginary relationships are so absurd as to be unworthy of our attention, yet they have gained the support of adherents of prominent schools of thought as well as of quacks and charlatans.

One famous theory which we may appropriately use to introduce this section is the work of the school of Italian criminologists headed by Lombroso. Lombroso, who devoted his life to the study of the criminal, tells us as follows how he first discovered that there is a criminal type:

"In 1870 I was carrying on for several months researches in the prisons and asylums of Pavia upon cadavers and living persons, in order to determine upon substantial differences between the insane and criminals without succeeding very well. Suddenly the morning of a gloomy day in December, I found in the skull of a brigand a very long series of atavistic anomalies, above all an enormous middle occipital fossa, and a hypertrophy of the vermis, analogous to those that are found in inferior invertebrates. At the sight of these strange anomalies, as a large plain appears under an inflamed horizon, the problem of the nature and of the origin of the criminal seemed to me resolved; the characters of primitive men and of inferior criminals must be reproduced in our time."*

This discovery in a single individual led to the rearing of a complete theory that criminals could be identified by certain "stigmata," said to be atavistic in nature. This theory has been carried to extreme lengths. Criminal characteristics, we are told (26) by the Lombroso school, have dark and thick *hair*, sometimes woolly in texture; a *skull* which may vary in five directions: (a) it may be rounded like a dome, or (b) depressed like a roof that is flat or low, or (c) its vault may be keel-shaped, or (d) it may be bulging, with a protuberance on one side, or both sides, or in front, or behind, or (e) it may have a sugar-loaf appearance; *eyebrows* that are beady or scanty; a *nose* that is defective and is frequently without a bony skeleton; *ears* that are long and thick; *skin* that is pale and wrinkled; *lips* that are cleft; *teeth* with the molars undeveloped, wisdom teeth absent, the canine teeth overdeveloped, etc.

"Finally, to select from a list of remaining characteristics, we must add that, according to various authorities, the male criminal has often the bust of a female and the female criminal the beard of a man, and that both male and female suffer from infantilism; that the criminal has an ape-like agility and a prehensile foot; that he is left-handed and ambidextrous, with his right hand smaller than his left, and his left foot smaller than his right; that he stammers and squints; that he sleeps soundly, tattoos his body, is given to the early use of tobacco, is sensitive to the weather, and is seldom seen to blush!"

Goring (26) in an extensive study in which he reports the results of anthropological measurements on 3,000 English criminals, sub-

*From a speech made by Lombroso at the Congress of Criminal Anthropology, held at Turin in 1906.

jected many of the important claims of the Lombroso school to the test of careful measurement and statistical treatment. After finding the correlation between various measurements of the head, face, and other physical characters and the nature of the crime, he concludes:

"It will be seen that ten only of the thirty-seven characteristics have correlations with nature of crime greater than .1, and that the correlations of the remaining twenty-seven are either insignificant, relatively to their probable errors, or so small in value as to be legitimately ignored in such limited samples as those we have been examining. Of the ten above .1 in value, three only are above .2; and only one above .3 in value. With the exception of these ten, which will require more detailed investigation, we may say that these physical characters have no significant association with the nature of the crime committed. In other words, we conclude that if there be any real association between physical characteristics and crime, this is so microscopic in amount as not to be revealed by the values of our correlation ratios and coefficients of contingency."

These ten characters which show correlation were studied with reference to their relation to other selective factors. His inquiry led him to state, "Physical differences exist between different kinds of criminals, precisely as they exist between different kinds of law-abiding people. But, when allowance is made for a certain range of probable variation, and when they are reduced to a common standard of age, stature, intelligence and class, etc., these differences tend entirely to disappear." Finally he compared the physical characteristics of criminals with those of non-criminals and decided, "No evidence has emerged confirming the existence of a physical criminal type, such as Lombroso and his disciples have described." "There is no such thing as a physical criminal type."

Similar systems of character analysis based upon the superficial observation of external characteristics have been advocated for use in industry and vocational guidance. One thoroughly worked-out system is that sponsored by Katherine Blackford (8). Her book on *The Right Job—How to Choose, Prepare for, and Succeed in It*, is most intriguing. Its appeal to the semi-educated man must be overwhelming. Here are diagrams and pictures showing in the clearest outline the significance of the shape of the head, the face and the facial profile, the shape, size and flexibility of the hand, bodily proportions, complexion, etc. The captions under the

pictures are also disarmingly convincing (8, p. 35): "Example of high, narrow head, showing ambition, keen intelligence, and intuition, with humor, spirit, sympathy, and affection." (8, p. 131.) "Professional type, with indications of mechanical and inventive ability. Note blonde color, height and width of forehead, especially in upper middle, with squareness of jaws and width between eyes."*

One may pick out absurd generalizations on every page (8, p. 193): "The person who finds the applause of others a heavy stimulant has a short upper lip." (8, p. 166.) "The organizer must have a 'high head,' both at the crown and also at the top or that portion which is just back of the hairline and above the temples."** and so on.

There is a kind of persuasiveness in much of the kind of reasoning used by those who claim to be able to read character from physiognomy. If the experiences of life leave their traces in the nervous system so as to effect future behavior, it is plausible that traces of experience are left in facial expression. It is well known that the emotions tend to heighten all muscular activity and result in various muscular tensions. So a life of strain leads to tensions around the mouth or in the furrow between the eyes that eventually should result in a set expression. Fatigue also leads one to try to relieve the fatigued parts by extra effort in unused muscles, and consequently the strain of fatigue is often imparted to facial expression.

The arguments underlying shape or contour of head or face are obscure. Most of them depend on analogies to function that have little basis in fact.

Several of the claims commonly put forward by these character readers have been put to the experimental test. One of the most ambitious of these is reported by Cleeton and Knight (12). Their work was carried out by obtaining physical measurements and character ratings on twenty-eight college students, members of two sororities and a fraternity. Eight character traits were selected for study,—sound judgment, intellectual capacity, frankness, will-power, ability to make friends, leadership, originality, and impul-

* From Blackford, Katherine M. H., and Newcomb, A., *The Right Job—How to Choose, Prepare for, and Succeed in It*, pp. 35 and 131. By permission of Katherine M. H. Blackford.

** *Ibid.*, pp. 193 and 166. By permission of Katherine M. H. Blackford.

518 Diagnosing Personality and Conduct

siveness, these being traits upon which the phrenologists agree best in their systems of diagnosis. A list of the physical attributes commonly alleged to be diagnostic of each of these traits was assembled. There were twenty-eight physical characteristics for judgment, twenty-nine for intelligence, thirty-six for will-power, thirteen for ability to make friends, etc. Accurate measurements with calipers, tapes, head square, and other instruments were made for each of the 201 physical characteristics mentioned.

The second part of the experiment consisted in obtaining ratings on the character traits. Each of the members of three fraternities rated all of the other members of his fraternity on each of the eight traits. These ratings, known as the ratings of "close associates," are shown to be highly reliable. Ratings by "casual observers" were also obtained by seating the subjects on a platform where they were rated by seventy judges—business men, school superintendents, and students of personnel management. These judges were accustomed to size people up in selecting for employment, but claimed they did not use any system. These ratings by the casual observers were also highly reliable, but they correlated low with the ratings by the close associates.

Finally, each of the ratings was correlated with each of the physical measurements chosen as being symptomatic of a given trait. These correlations as averaged are presented in the following table:

TABLE 87

CORRELATION BETWEEN RATINGS OF CHARACTER FACTORS AND PHYSIOGNOMIC MEASUREMENTS ALLEGED TO BE SYMPTOMATIC OF THESE FACTORS

(from Cleeton and Knight, 12, p. 229)

<i>Traits</i>	<i>Ratings of close associ- ates and physical mea- sures</i>	<i>Ratings of casual ob- servers and physical mea- sures</i>	<i>Ratings of close associ- ates and those of casual observers</i>
Judgment	— .01	.14	.32
Intelligence03	.05	.02
Frankness05	.15	.21
Friendliness	— .11	.19	.18
Will-power	— .07	.04	.26
Leadership	— .04	.07	.31
Originality09	.08	.32
Impulsiveness10	— .07	.20

The investigators conclude that "physical factors purporting to measure the same trait do not present even a suspicion of agreement"; that "the correlation between ratings of casual observers and physical measurements is best represented by 0.000"; that "physical measurements which underlie character analysis agree neither with themselves nor with other measures of character." (12, p. 230.)

These sweeping conclusions do not seem entirely justified by the correlations reported. It is quite possible that the *average* of the twenty-eight physical measurements alleged to indicate judgment failed to do so, but that some one or more of the measures may have some significance. There may perhaps be a kernel of truth in some of the physiognomist's assertions. For instance, in this study the correlation between ratings on *judgment* by close associates and the ratio of anterior length of head to posterior length of head is $.29 \pm .12$, the ratio of anterior inferior length of head to posterior central length is $.38 \pm .12$. To be sure, these are within the possible range of chance deviations from a zero correlation, but they are at least suggestive of a possible relationship. Cleeton and Knight's study, although it denies the possibility of any widespread truth in the assertions of character readers, does not destroy the possibility that out of the welter of claims there may be established some physical indicia of value in the diagnosis of character. It is deplorable that these investigators did not report all of their results in detail instead of assuming that the averages tell the whole story.

Among the claims advanced by the Blackford school of character reading is one concerning the significance of the facial profile. The convex type with prominent nose, sloping forehead, and receding chin is said to indicate a person possessing "superabundance of energy," and one who is "keen, alert, quick, eager, aggressive, impatient, positive, and penetrating." The concave type of profile with prominent chin and high forehead indicates a person who is characterized by mildness and who is "slow of thought, slow of action, patient in disposition, plodding." Alice L. Evans (32, 19), under direction of Clark Hull, investigated the character significance of the profile, using a specially devised measuring instrument for obtaining an accurate index of facial convexity. Twenty-five university women, members of the same sorority, rated each other on optimism, activity, ambition, will-power,

520 Diagnosing Personality and Conduct

domination, popularity, blondness. These ratings were highly reliable with the exception of those of popularity. The significant relationships in the study are shown in the following table:

TABLE 88
CORRELATIONS BETWEEN MEASUREMENTS OF FACIAL PROFILE AND PERSONALITY RATINGS *

(from Evans and Hull, 32, p. 130)

	<i>Optimism</i>	<i>Activity</i>	<i>Ambition</i>	<i>Will-power</i>	<i>Domination</i>	<i>Popularity</i>
Convexity—whole face						
with nose	+ .10	— .05	— .17	— .13	— .11	— .03
Convexity—chin to eye-						
brow without nose ..	+ .37	+ .39	+ .33	+ .34	+ .24	+ .17
Height of forehead from						
eyebrow to hairline...	— .17	— .29	— .23	— .39	— .22	— .10

Hull's conclusions from Evans' work are helpful (32, p. 129):

"For the most part, convexity also seems unrelated to the various character traits. A possible exception to this seems to be convexity of the lower face (chin to eyebrow without nose), which yields a curiously consistent series of relatively large positive correlations with all of the character traits. Because of the small number of subjects the probable error of these coefficients is rather large, but the correlations are nevertheless suggestive. A second possible exception lies in the series of negative correlations between the various traits and the height of the forehead from eyebrow to hairline. If these coefficients may be taken at face value, they indicate that *low* foreheads tend to indicate optimism, activity, will, etc.; but here again we must reserve judgment because of the small number of subjects. These coefficients are sufficiently suggestive to warrant further investigations." **

The possibility that some of these relationships may be greater than chance is tantalizing. If they should so prove it might be surmised that the judges unwittingly employed these physiognomic characteristics as a basis for their judgments. There is a strong presumption, however, for believing that the relationship between objective measures of conduct and physiognomic traits is purely a chance relationship.

* From Hull's *Aptitude Testing*. Copyright 1928 by World Book Company, Publishers, Yonkers-on-Hudson, New York.

** *Ibid.*

Estimating Character from Photographs

It seems probable that, if it is not possible to estimate character from direct face-to-face observation on the basis of physiognomic characters, there is little possibility that one can judge character from photographs. However, there are systems of character judgment based on photographs, and their claims must be considered. Two studies have been made on the relationship of character traits and judgments of photographs, one by Miss Cogan, working under the direction of H. L. Hollingworth (30), and the other by Miss McCabe, working under Clark Hull (32, 45). In each case a group of college women rated each other on character traits. Since each group was composed of members of a sorority, the subjects knew each other intimately, and their judgments of character differences were as accurate as could be obtained under any circumstances. The combined ratings had very high reliability. A second squad of judges, unacquainted with the first, rated photographs of the subjects for the same character traits as those used by the subject groups. The correlations between the ratings made by intimate associates on the basis of acquaintance and those made from examination of photographs are given in the following table:

TABLE 89
CORRELATION BETWEEN REAL CHARACTER TRAITS AND JUDGMENTS OF CHARACTER
BASED ON PHOTOGRAPHS *
(from Hull)

<i>Trait</i>	<i>McCabe's results</i>	<i>Cogan's results</i>
Neatness	+ .07	+ .05
Conceit	+ .21	+ .19
Sociability	+ .12	+ .29
Humor	- .07	+ .33
Likability	+ .15	+ .38
Intelligence	+ .40	+ .51
Refinement	+ .17	+ .51
Beauty	+ .61	+ .55
Snobbishness	+ .32	+ .56
Vulgarity	+ .10	+ .65

McCabe's results are the more dependable since they are based on a larger number of cases (forty). Only two of the cor-

* From Hull's *Aptitude Testing*. Copyright 1928 by World Book Company, Publishers, Yonkers-on-Hudson, New York.

relations are of any size—those for beauty (+.61) and for intelligence (+.40). We would expect that beauty could be judged from a photograph, for beauty is “only skin deep.” As for the correlation of .40 with intelligence, this disappears when checked up against objective evidence of intellectual ability. It is probable that the relationship of .40 holds because all the judges used certain features such as seriousness, wearing glasses, a bright eye, etc., as aids in estimating intelligence from the photographs. Estimations of character from a photograph are worthless both because a single judgment is very unreliable and also because there is probably no relationship between physiognomic features and character traits.

Blond-Brunette Types

Another of Blackford's claims that has been investigated experimentally concerns blond and brunette types. There is a widespread belief that blonds and brunettes possess certain well-defined characteristics. Blackford has been more definite than others, as the following description shows (7, p. 144):

“In brief, always and everywhere, the normal blond has positive, dynamic, driving, aggressive, domineering, impatient, active, quick, hopeful, speculative, changeable, and variety-loving characteristics; while the normal brunette has negative, static, conservative, imitative, submissive, cautious, painstaking, patient, plodding, slow, deliberate, serious, thoughtful, specializing characteristics.”

These descriptions certainly challenge inquiry, and the honors for a critical and clear-cut investigation go to Paterson and Ludgate (52). The twenty-six characteristics mentioned by Blackford were listed in random order on a rating sheet. The directions for rating follow:

“Select from among all the people you know *very well* two who are pronounced blonds and two who are pronounced brunettes, and rate them with respect to the characteristics listed below. Put either a plus (+) or a minus (—) sign after *each* characteristic for each of the four persons you select for rating. If your first blond is positive, put a plus sign after that characteristic, if not, put a minus sign and so on for the rest of the characteristics.”

External Signs of Conduct

523

Ninety-four college students acted as judges. If Blackford's claims are correct, there should be a preponderance of plus signs for blonds and negative signs for brunettes for the trait "positive," and so on down through the list. As a matter of fact the differences between blonds and brunettes are remarkably small, as is shown in the accompanying table.

TABLE 90
PERCENTAGES OF BLONDS AND BRUNETTES RATED AS POSSESSING CERTAIN
PERSONALITY TRAITS
(from Paterson and Ludgate, 52, p. 125)

<i>Blond traits</i>	PERCENTAGE RATED PLUS	
	<i>187 blonds</i>	<i>187 brunettes</i>
Positive	81	84
Dynamic	63	64
Driving	49	50
Aggressive	62	56
Domineering	36	36
Impatient	56	51
Active	88	82
Quick	70	68
Hopeful	85	85
Speculative	53	51
Changeable	53	43
Variety-loving	66	62

<i>Brunette traits</i>	PERCENTAGE RATED PLUS	
	<i>187 blonds</i>	<i>187 brunettes</i>
Negative	16	17
Static	28	31
Conservative	51	61
Imitative	39	40
Submissive	25	26
Cautious	54	60
Painstaking	56	61
Patient	43	52
Plodding	27	31
Slow	20	24
Deliberate	47	57
Serious	58	72
Thoughtful	67	70
Specializing	52	45

These results show not only that there are no differences between blonds and brunettes of the kind claimed, but that there is not even generally prevalent an organized or systematic belief concerning differences between blonds and brunettes which might have influenced the ratings.

Shape of the Hand

Still another school of character readers professes to be able to use the shape and size of the hand for describing character traits. As with the other systems there is just enough analogy between the use of the hands in different occupations and the shape best adapted to this use to make the claims seem plausible. We read that the longer the first finger is when compared to the second, the more ambitious a person is; the farther a person can bend his fingers backward, the keener his mind; the longer the fingers in proportion to the length of the palm, the stronger the tendency to impulsiveness. These claims are so absurd as to be beneath the dignity of serious consideration, but since they attract the attention of business men and others who must deal with personnel problems, their emptiness should be revealed.

Fortunately Miss MacLaurin (32, 44), another of Clark Hull's students, has carried on an investigation to determine the facts. Again thirty members of a sorority rated each other on character traits mentioned in the literature of *chirognomy*, and the experimenter made careful measurements of their hands. The following table illustrates the findings:

TABLE 91
CORRELATIONS BETWEEN CERTAIN MEASURED PROPORTIONS OF THE HAND AND
THE CHARACTER TRAITS ALLEGED TO BE INDICATED BY EACH
(from Hull*)

Hand trait	Character trait	Correlation P.E.
Difference in height between first and second finger	Ambition	+ .19 ± .12
Extent of backward flexion of the hand, measured in degrees	Keeness of mind	+ .13 ± .12
Degree of taper of fingers	Refined sensibility	+ .16 ± .12
Length of fingers divided by length of palm	Impulsiveness ...	+ .29 ± .11
Length of thumb divided by length of palm and second finger	Force of character	— .23 ± .12

Here again the correlations are suggestive, but the probable errors are so large as to prevent one from concluding that anything but a chance relationship exists.

* From Hull's *Aptitude Testing*. Copyright 1928 by World Book Company. Publishers, Yonkers-on-Hudson, New York.

Handwriting

The last in our list of external signs used as a basis for judging character is handwriting. Handwriting differs from the preceding signs in being a product of activity rather than a characteristic of the body itself. It is quite conceivable that one's temperamental qualities should express themselves in handwriting, and handwriting is fortunately left as a permanent record to be studied, facilitating methods by which this hypothesis can be tested. Many persons have been fascinated by the hypothesis that handwriting reveals character. Observation of handwriting is to-day one of the common methods used to obtain a quick and easy diagnosis of character. It must be confessed that certain persons actually have acquired a skill in the art of reading handwriting that enables them to diagnose traits with an accuracy greater than the expectations of chance. But this art will not be generally useful until it can be codified into definite rules which will stand under the scrutiny of experimental investigation.

Clark Hull (33), who has conducted one such investigation, has reviewed for us some of the history of character reading through handwriting. Camillo Baldo published a treatise on the subject in 1662, and the matter attracted the attention of such eminent men as Leibnitz, Goethe, and Sir Walter Scott. An extensive body of literature has grown up on the subject. That men like William Preyer (58), professor of physiology at Jena, and George Schneidemühl (65), professor of comparative pathology at the University of Kiel, are known as graphologists shows the extent to which the "science" of reading handwriting has won a recognized place in academic circles. Most of the graphologists, however, build systems and make claims without bringing them under experimental scrutiny.

One of the most elaborate systems of graphology is that of Robert Saudek (63), a man who has made an intensive study of handwriting but has mixed his physiological studies with vague speculations. This system resolves itself into such generalizations as:

"Predominance of the lower projection points to a precise, healthy, natural functioning of the muscular tonus, and often, in addition, a practical mind and commercial sense" (p. 298).

"The lack of vertical readjustments of the writing surface points to indolence and lack of adaptability" (p. 294).

526. Diagnosing Personality and Conduct

In describing his method of detecting *dishonesty*, Saudek mentions ten characteristics of handwriting. Number 10 illustrates the method.

"10. The letters *o*, *a*, *d*, *g*, *q* are open at the base and are therefore (with the exception of the *o*) written in at least two strokes, with a clockwise movement." [He adds] "We know that this feature is the only reliable feature (as regards the formation of the letters) of the French school of graphology, and also that in the files of the police it occurs in 30 per cent of the signatures alone of habitual thieves of both sexes" (p. 286).

Evidence presented by Saudek is quite unconvincing, even in its fundamental assumptions, betraying a lack of knowledge of the ways in which character is organized. For instance, in demonstrating his ability to detect dishonesty from handwriting, Saudek collected specimens of seventy-three individuals from nineteen business firms. In fourteen cases he diagnosed dishonesty, and in fifty-nine cases honesty. The firms in question confirmed the accuracy of his diagnosis in all these fourteen cases, for in each of these cases the writer had either been convicted in the courts of embezzlement, or had confessed his guilt but had not been prosecuted. This dramatic testimony to the power of graphology may be convincing to the business man, but the logic of the argument hardly squares with other facts that are known concerning the specificity of conduct.

Binet (5), the famous French psychologist, was the first to study graphology scientifically. He set out to study the possibility of determining sex from handwriting. To each of his subjects he presented envelopes, 180 in all, addressed by an equal number of men and women. Those who had no special training were able to make from 66 to 73 per cent (average 69 per cent) correct guesses, where a series of guesses which are due to chance alone would yield about 50 per cent correct. One graphologist made a score of 75 per cent, and the famous graphologist Crépieux-Jamin (13) made a score of 79 per cent. Downey conducted a similar experiment. Using untrained subjects, percentages of correct judgment running from 60 to 77.5, with an average at 67.3, were obtained.

In another experiment Binet secured the handwriting of eleven notorious assassins and paired each with a sample of handwriting by a law-abiding man. Again the famous graphologist

Crépieux-Jamin exceeded the chance expectation in making only three errors out of the eleven judgments, or a score of 73 per cent. Two other graphologists were able to make only six out of eleven correct judgments each.

Hull and Montgomery (33) studied the claims of graphology by comparing exact measures of handwriting characteristics with ratings of the character traits they were supposed to represent in the same individual. First these investigators made a careful survey of the literature of graphology and culled out the assertions about which there seems to be the most agreement. These are given in the following table.

TABLE 92

CLAIMS MADE OF RELATIONSHIP BETWEEN CHARACTER TRAITS AND
HANDWRITING CHARACTERISTICS

(from Hull and Montgomery, 33, p. 66)

<i>Character trait</i>	<i>Indicated by handwriting having the following characteristics</i>
Ambition	Lines of writing sloping upward
Pride	Lines of writing sloping upward
Bashfulness	Writing traced with heavy lines
Force	(a) Heavy lines; (b) heavy bars on the <i>r</i> 's
Perseverance	Long bars on the <i>r</i> 's
Reserve	Closed <i>a</i> 's and <i>o</i> 's

Seventeen students, members of a fraternity, submitted samples of their handwriting. These samples were written in a uniform but natural way, each copying a chosen passage. Exact measures were made of the slope of the lines, the thickness of the lines, the thickness of the bars on the *r*'s and also their length, and count was made of the *a*'s and *o*'s which were closed. At the same time the subjects ranked each other on the character traits in question. Correlations computed between the character traits and the handwriting characteristics alleged to be symptomatic of them are given in Table 93.

The average of these correlations is $-.016$, which represents about the amount of assurance that one should give to the claims of graphologists. Any one of the correlations in the table could have occurred as a chance deviation from zero, although certain of the coefficients are high enough to warrant further investigation. It would not be surprising to find that forceful people write heavier than timid people, or that bashful people write lightly,

528 Diagnosing Personality and Conduct

TABLE 93

CORRELATIONS OF RATINGS OF CHARACTER AND HANDWRITING (from Hull and Montgomery)

Ambition with upward sloping lines	-.20
Pride with upward sloping lines	-.07
Bashfulness with fineness of lines	-.45
Bashfulness with lateral narrowness of <i>n</i> 's and <i>u</i> 's	+.38
Force with heavy handwriting	-.17
Force with heavy bars on <i>t</i> 's	-.06
Perseverance with length of bars on <i>t</i> 's00
Perseverance with length of bars on <i>t</i> 's, varying size of writing compensated for	+.16
Reserve with closed <i>a</i> 's and <i>o</i> 's	-.02

but such does not seem to be the case when these characteristics of people are actually compared with their handwriting.

Another of Hull's students, Miss L. E. Brown (9), starting out with a strong belief in graphology, performed a similar experiment with similar results as shown in the following table.

TABLE 94

CORRELATION BETWEEN RATINGS OF CHARACTER TRAITS AND HANDWRITING* (from Hull, 32, p. 150)

<i>Character trait</i>	<i>Handwriting trait</i>	<i>Correlation</i>
Bashfulness	Width of down stroke	+.11
Ambition	Tendency to upward slope as line crosses page	+.23
Persistence	Width of down strokes	-.05
Persistence	Disconnected writing—per cent of breaks of line within words	-.03
Personal neatness	Neatness in appearance of writing	+.23
Personal individuality..	Individuality in appearance of writing..	+.15

Hull suggests the possibility that the relation between neatness and individuality and the corresponding characteristics in handwriting may represent genuine tendencies.

Summary

Out of this mass of assertions and claims for systems of character reading the positive findings of exact measurement are pitifully few. One can say that they reduce themselves to (a) tall, large men tend to be more aggressive and sociable and to make

*From Hull's *Apptude Testing*. Copyright 1928 by World Book Company, Publishers, Yonkers-on-Hudson, New York.

more successful leaders than short men; (b) there are temperamental differences between fat men with large trunks and small limbs and thin men with small trunks and larger, more athletic limbs; the former tend to be more open, friendly, jolly, and sociable; the latter more reserved, retiring and unsociable; (c) personal beauty may (naturally) be determined from photographs; (d) sex may be determined from handwriting. It is, of course, understood that in none of these is the relationship more than suggestive, nor can the diagnosis from these characteristics be made with much more than chance accuracy. There are also vague indications supporting the following small yet possibly real relationships: (a) certain ratios of head measurements may have a very slight relationship to intelligence or personality qualities; (b) neatness, individuality, and similar traits may have some relationship to the expression of these traits in handwriting.

Beyond these meager findings it is safe to conclude that one cannot use external characteristics or signs on the body to diagnose conduct. Conduct is a dynamic relationship between situation and response, so that what conduct may be expected from a person is a secret hidden in the nervous system, only to be discovered by the actual reactions a person makes.

Blackford and Newcomb claim in one of their books (7, pp. v, vi):

"Within the six years since the first publication of this book we have seen the plan adopted in its essentials by practically all progressive employers of large numbers of men and women in this country and also by many in other lands. We have seen hundreds of investigators, engineers, students, personnel experts and university research workers studying the plan, experimenting with it, collecting data in regard to it and elaborating upon it. We have seen the United States Government adopt it for all industries working on government contracts, and conducting free courses of instruction in a number of prominent universities teaching hundreds of men and women how to operate the plan in industry."

That such a claim must be exaggerated is obvious. However, Kornhauser (41) in 1922 made a check on the practice in industry by sending out a questionnaire to 100 employment managers in industrial plants and to 100 insurance agency managers asking what was their practice in using a system of character

530 Diagnosing Personality and Conduct

analysis, if any. Forty-three and twenty-two replies, respectively, were received. Four industrial plants and two insurance agencies, a total of six, used such a system. Three used Blackford's system, two used a combination of Blackford's system with some other, and one did not specify the system. While it is evident that Blackford's claims are exaggerated, we can feel some concern that even six out of 200 should be deceived by claims which do not stand up in the light of experimental analysis and which consequently have no scientific standing.

Although the evidence is overwhelmingly against the possibility of diagnostic significance in these external signs, there is need for further experimental study. Most of the experiments described in this chapter have employed a small number of subjects, with the result that some of the correlations reported seem suggestive, although the possibility that they are chance deviations from a zero relationship is not removed. These experiments should be repeated using larger numbers of subjects. Again it is likely that some of these correlations have been produced by the fact that in making judgments of character the judges have used as a basis for their ratings precisely the physical characteristics whose significance they were testing. The criteria for crucial experiments ought to be objective enough to provide a real check.

Let us close this section by again pointing out the brazenness and shamelessness with which the charlatans, and there are many of them, push forward their claims. They usually have something to sell, and they make their living by selling it. Their assertions are dressed up in the most attractive way. They play upon the common foibles and weaknesses of mankind, who live in hopes and who are seeking praise and encouragement. Many of the successes of these impostors in making correct diagnoses are due not to an application of their own systems but to a keen evaluation of the characteristics of those whom they face.

REFERENCES

1. ADAMS, H. F., "Psychological Goldbricks," *Scribner's Magazine*, 70: 94-101 (1921).
2. ———, "The Mythology and Science of Character Analysis," *Scribner's Magazine*, 69: 569-575 (1921).
3. BELLINGRATH, G. C., *Qualities Associated with Leadership in the Extra-curricular Activities of the High School*, Teachers College Contributions to Education, No. 399 (1930).

4. BENHAM, W. G., *The Laws of Scientific Hand Reading* (G. P. Putnam's Sons).
5. BINET, A., *Les Révélation de L'Écriture d'après un Contrôle Scientifique* (Paris, Alcan, 1906).
6. BLACKFORD, K. M. H., *Reading Character at Sight* (New York Independent Corporation, 1918).
7. BLACKFORD, K. M. H., and NEWCOMB, A., *The Job, The Man, The Boss* (Doubleday, Page & Co., 1914, 1921, 1925).
8. BLACKFORD, K. M. H., and NEWCOMB, A., *The Right Job—How to Choose, Prepare for, and Succeed in It*, A Treatise for Parents, Guardians, Teachers, and Vocational Counselors (The Review of Reviews Corporation, 1924).
9. BROWN, L. E., *An Experimental Investigation of the Alleged Relations between Certain Character Traits and Handwriting*, A. B. thesis, University of Wisconsin, 1921.
10. BURTT, H. E., *Principles of Employment Psychology* (Houghton Mifflin Company, 1926).
11. CLEETON, G. U., "Estimating Human Character," *Scientific Monthly*, 23: 427-431 (1926).
12. CLEETON, G. U., and KNIGHT, F. B., "Validity of Character Judgments Based on External Criteria," *Journal of Applied Psychology*, 8: 215-231 (1924).
13. CRÉPIEUX-JAMIN, J., *L'Écriture et le caractère* (Paris, Alcan, 1895).
14. DE GIOVANNI, A., *Morfologia del Corpo Umano* (Milano, 1891).
15. ———, "Studi Morfologici o Contribuzione della Clinica e della Dottrina dei Temperamenti," *Osservatore*, 16 (1880).
16. DOWNEY, J., "Judgments in the Sex of Handwriting," *Psychological Review*, 17: 205-216 (1910).
17. DUNLAP, K., "Fact and Fable in Character Analysis," *Annals of American Academy of Political and Social Science*, 110: 74-80, No. 199 (1923).
18. ———, "The Reading of Character from External Signs," *Scientific Monthly*, 15: 153-165 (1922).
19. EVANS, A. L., *The Alleged Relations between the Face and the Character*, A. B. thesis, University of Wisconsin, 1921.
20. FARR, C. B., "Bodily Structure, Personality, and Reaction Types," *American Journal of Psychiatry*, 7: 231-244 (1927, 1928).
21. FOSBROOKE, G. E., *Character Qualities Outlined and Related* (G. P. Putnam's Sons, 1922).
22. ———, *Character Reading through Analysis of Features* (G. P. Putnam's Sons, 1914).
23. FRENCH, W. L., *The Psychology of Handwriting* (G. P. Putnam's Sons, 1922).

532 Diagnosing Personality and Conduct

24. FRENCH, W. L., *What Your Handwriting Reveals* (Spencerian Pen Co., 1915).
25. GARRETT, H. E., and KELLOGG, W. N., "The Relation of Physical Constitution to General Intelligence, Social Intelligence, and Emotional Stability," *Journal of Experimental Psychology*, 11: 113-129 (1928).
26. GORING, C., *The English Convict—A Statistical Study* (London, His Majesty's Stationery Office, 1913).
27. GOWIN, E. B., *The Executive and His Control of Men* (The Macmillan Company, 1915).
28. GRUHLE, H. W., "Der Körperbau der Normalen," *Arch. f. Psychiatr. u. Nervenkr.*, 77: 1-32 (1926).
29. HEIDREDER, E., "The Normal Inferiority Complex," *Journal of Abnormal and Social Psychology*, 22: 243-258 (1927).
30. HOLLINGWORTH, H. L., *Vocational Psychology and Character Analysis* (D. Appleton and Company, 1929).
31. HORST, L. v. D., "Experimentell-psychologische Untersuchungen zu Kretschmer's 'Körperbau und Charakter,'" *Zsch. f. d. ges. Neur. u. Psychiatr.*, 93: 341-380 (1924).
32. HULL, C. L., *Aptitude Testing* (World Book Company, 1928).
33. HULL, C. L., and MONTGOMERY, R. B., "An Experimental Investigation of Certain Alleged Relations between Character and Handwriting," *Psychological Review*, 26: 63-75 (1919).
34. JASTROW, J., "The Antecedents of the Study of Character and Temperament," *Popular Science Monthly*, 86: 590-613 (June, 1915).
35. KENAGY, H. G., "Do Blonds Make the Best Salesmen?" *Sales Management*, 5: 325-326 (Feb., 1923).
36. KENAGY, H. G., and YOAKUM, C. S., *The Selection and Training of Salesmen* (McGraw-Hill Book Company, 1925).
37. KITSON, H. D., *The Psychology of Vocational Adjustment* (J. B. Lippincott Company, 1925).
38. KLAGES, L., *Science of Character* (London, G. Allen & Unwin, 1929).
39. KOLLE, K., "Der Körperbau der Schizophrenen," *Arch. f. Psychiatr. u. Nervenkr.*, 72: 40-88 (1924). Körperbauuntersuchungen an Schizophrenen," *ibid.*, 75: 21-61 (1925).
40. ———, "Klinische Beiträge zum Konstitutionsproblem," *Arch. f. Psychiatr. u. Nervenkr.*, 77: 183-238 (1925); 78: 93-165 (1926).
41. KORNHAUSER, A. W., and JACKSON, A. W., "A Note on the Extent to Which Systems of Character Analysis Are Used in the Business World," *Journal of Applied Psychology*, 6: 302 (1922).
42. KRETSCHMER, E., *Physique and Character* (Harcourt, Brace and Company, 1925).

43. LOMBROSO, C., *Crime, Its Causes and Remedies* (Little, Brown and Company, 1911).
44. MACLAURIN, D. D., *An Experimental Investigation of Certain Alleged Relations between Physical Characteristics of the Hand and Mental Traits*, A. B. thesis, University of Wisconsin, 1921.
45. McCABE, F. E., *The Relation between Character Traits and Judgments of Character Based on Photographs*, A. B. thesis, University of Wisconsin, 1926.
46. MAUZ, F., "Ueber Schizophrene mit pyknischen Körperbau," *Zsch. f. d. ges. Neur. u. Psychiatr.*, 86: 96-122 (1923).
47. MOHR, G. J., and GUNDLACH, R. H., "The Relation between Physique and Performance," *Journal of Experimental Psychology*, 10: 117-157 (1927).
48. NACCARATI, S., and GARRETT, H. E., "The Influence of Constitutional Factors on Behavior," *Journal of Experimental Psychology*, 6: 455-465 (1923).
49. ——— and ———, "The Relation of Morphology to Temperament," *Journal of Abnormal and Social Psychology*, 19: 254-263 (1924).
50. NACCARATI, S., "The Morphologic Basis of the Psychoneuroses," *American Journal of Psychiatry*, 3: 527-545 (1924).
51. OLIVIER, H. G., "Der Körperbau der Schizophrenen," *Zsch. f. d. ges. Neur. u. Psychiatr.*, 80: 489-498 (1922).
52. PATERSON, D. G., and LUDGATE, K., "Blonde and Brunette Traits," *Journal of Personnel Research*, 1: 122-127 (1922).
53. PATERSON, D. G., "A Note on Popular Pseudo-psychological Beliefs," *Journal of Applied Psychology*, 7: 101-102 (1923).
54. ———, "Personality and Physique," in *The Measurement of Man* (The University of Minnesota Press, 1930).
55. ———, *Physique and Intellect*, in *The Century Psychology Series* (The Century Co., 1930), 304 pp.
56. PENDE, N., *Constitutional Inadequacies: An Introduction to the Study of Abnormal Constitutions* (Lea & Febiger, 1928).
57. POPE, D. V., "The Interpretation of the Human Face from Photographs," *Bulletin of Randolph-Macon Women's College*, Vol. VIII (1922).
58. PREYER, W., *Zur Psychologie des Schreibens* (Hamburg, 1895).
59. RAVA', G., "Costituzione Neurastenica e Costituzione Distimica," *Giornale di Psichiatria Clinica e Tecnica Manicomiale*, 49: 3-4 (1921); "Istinto e Emizione," *ibid.*, 1-2 (1922).
60. REXFORD, J., *What Handwriting Indicates* (G. P. Putnam's Sons, 1904).
61. ROHDEN, F. V., "Körperbauuntersuchungen an geisteskranken und gesunden Verbrechern," *Arch. f. Psychiatr. u. Nervenkr.*, 77: 151-164 (1924).

534 Diagnosing Personality and Conduct

62. ROHDEN, F. V., and GRÜNDLER, W., "Über Körperbau und Psychose," *Zsch. f. d. ges. Neur. u. Psychiat.*, 95: 37-78 (1925).
63. SAUDEK, R., *Experiments with Handwriting* (George Allen & Unwin, London, 1928; New York, William Morrow and Company, Inc., 1928).
64. ———, *The Psychology of Handwriting* (London, George Allen and Unwin, 1928; New York, George H. Doran Company, 1926).
65. SCHNEIDEMUHL, G., *Handschrift und Charakter* (Leipzig, 1911).
66. SHAW, F. C., "A Morphologic Study of the Functional Psychoses," *State Hospital Quarterly*, 10: 413 (1924-1925).
67. SHELDON, W. H., "Ability and Facial Measurements," *Personnel Journal*, 6: 102-113 (1927).
68. ———, "Morphologic Types and Mental Ability," *Journal of Personnel Research*, 5: 447-451 (1927).
69. ———, "Social Traits and Morphologic Types," *Personnel Journal*, 6: 47-55 (1927).
70. SHERMAN, E. B., *An Experimental Investigation Concerning Possible Correlation between Certain Head Measurements and University Grades*, A. B. thesis, University of Wisconsin, 1923.
71. SNOW, A. J., "Can We Read Human Character?" *Hygiene*, 3: 185-188 (April, 1925).
72. VIOLA, G., "Studi di Morfologia Clinica" (in De Giovanni, A.) *Lavori dell'Istituto di Clinica Medica di Padova*, Vol. 2, Milano, 1905.
73. ———, "Nuovi Studi di Morfologia Clinica," in De Giovanni, A., *Lavori dell'Istituto di Clinica Medica di Padova*, Vol. IV (Milano, 1909).
74. WEISSENFELD, F., "Beiträge zum Problem Körperbau und Charakter," *Zsch. f. d. ges. Neurol. u. Psychiat.*, 96: 173-208 (1928).
75. WERTHEIMER, F. I., "Constitutional Factors as an Aid to the Early Evaluation of Behavior Disorders," *Journal of the American Medical Association*, 88: 22-24 (1927).
76. WERTHEIMER, F. I., and HESKETH, F. E., "A Historical Review," *Medicine*, 5: 375-461 (1926).
77. ——— and ———, "A Minimum Scheme for the Study of the Morphologic Constitution in Psychiatry, with Remarks on Anthropometric Technique," *Archives of Neurology and Psychiatry*, 17: 93 (1925).
78. ——— and ———, "Observation and Remarks on the Physical Constitution of Female Psychotic Patients," *American Journal of Psychiatry*, 6: 499-506 (1927).
79. ——— and ———, "The Significance of the

External Signs of Conduct

535

Physical Constitution in Mental Disease," *Medicine*, 5: 375-467 (1926).

80. WERTHEIMER, F. I., and HESKETH, F. E., *The Significance of the Physical Constitution in Mental Disease* (The Williams and Wilkins Company, 1926).

Chapter XV

MEASURES OF THE ENVIRONMENT

UP to this point the measures which have already been described in this book have to do with *responses* of one sort or another. Now it remains to point out that in investigations concerning the causes and development of conduct it is often desirable to have knowledge concerning the *situations* in which conduct takes place. The situations are as manifold as the responses, and it is as impossible to describe succinctly all of the permutations and combinations of the environment as it is to describe all of the vagaries of conduct. A first step in the problem of describing environmental differences is to consider the general richness or adequacy or level of the child's environment. The very statement of the problem makes it one of measurement, for differences in richness or adequacy of the environment presuppose measurement. To describe all existing gradations there must be a scale running from the very poor, meager and insufficient environment to the rich, full, and adequate environment.

It will quickly be seen that it is not possible to measure at once all phases of the environment; i.e., community, church, playground, school, and home. Most attention in this problem has been devoted to the home, for it is there that conduct has its source, and beyond question the social and economic level of the home is a prime factor in determining conduct. Such items as father's occupation, family income, home furnishings, and possessions have been suggested as possible measures of this vague thing which we call adequacy of the environment. As is usual in the problem of measuring human affairs, a search has been made for that item or those items which correlate most highly with the more general and abstract thing which we want to measure. In the following sections will be described attempts to measure (a) occupation level, (b) home background, and (c) certain combinations of the two, and (d) attempts to get at the

cultural background through the cultural responses of the children.

Measures of Occupational Level

The classification of occupations has been a troublesome and vexing matter. Decennially the federal census wrestles with this problem and at each succeeding census manages by referring to past experience to make the occupation classification more and more adequate. The census emphasizes the fact that it attempts to classify along *occupational* rather than *industrial* lines, and in the schedule for enumerators an entry for the name of the occupation or type of work comes before the entry for the industry. The census makes the following broad classification of occupations under nine heads:

- Agriculture, forestry, and animal husbandry
- Extraction of minerals
- Manufacturing and mechanical industries
- Transportation
- Trade
- Public service (not elsewhere classified)
- Professional service
- Domestic and personal service
- Clerical occupations

The Bureau of the Census publishes an *Index of Occupations—Alphabetical and Classified*. The index which was used in the fourteenth (1920) census contains 572 main occupations and occupational groups.

It must be evident at once that the census classification does not satisfy the needs of those who are seeking a scale by which to measure the level of occupation. For instance, under the heading "Manufacturing and mechanical industries" are to be found such diverse groups as "Manufacturers and officials" and "Laborers." The census classification, capable of indefinite subdivision, is useful for some purposes, but does not satisfy the requirements of a scale for measurement of social or economic level.

A rough scale was projected by Taussig in his *Principles of Economics*. He describes five non-competing groups—"non-competing in the sense that those born or placed in a given grade or group usually remain there, and do not compete with those

in other groups." We will use Taussig's phraseology in describing his five groups* (24, Vol. II, pp. 134-137):

"In the lowest group belong the day laborers, so-called: the diggers and delvers who have nothing to offer but their bodily strength. . . . In the next group belong those who, while not needing specialized skill, yet bear some responsibility, and must have some alertness of mind. Such, for example, are motormen on street railways. . . . In the third group belong the aristocracy of the manual laboring class: the skilled workmen. Such are carpenters, bricklayers, plumbers, machinists; the whole range of occupations where there is need for a sure eye, familiarity with tools, a deft and trained hand. . . . Next comes the group that approaches the well-to-do, the lower middle class, which avoids rough and dirty work, and aims at some sort of clerical or semi-intellectual occupation. Here are clerks, bookkeepers, salesmen, small tradesmen, railway conductors, firemen, superintendents, teachers of the lower grades. . . . Finally, we reach the class of the well-to-do; those who regard themselves as the highest class, and certainly are the most favored class. Here are the professions, so-called,—the lawyers, physicians, clergymen; teachers of the higher grades; salaried officials, public and private, in positions of responsibility and power; not least, the class of business men and managers of industry, who form in democratic communities the backbone of the whole group."

This rough scale serves as a very satisfactory beginning to later more detailed scales.

Counts (12), in his study of *The Selective Character of American Secondary Education*, states that the Taussig scale is difficult to use. He found that the lines between the groups were not clearly defined in actual industry. Furthermore, because of the meagerness of the information which he obtained from high school students by asking them the questions, "Father's occupation? Where or for whom does he work? Is he either owner or part owner of the business in which he works?" He pronounced the Taussig scale "unworkable without resorting to many arbitrary decisions." The classification adopted by Counts uses the census classification as a basis, aiming at the same time to get classes "of reasonable homogeneity from the standpoint of social status, position in the economic order, and intellectual outlook." It recognizes the following groups:

* From Taussig, F. W., *Principles of Economics*. By permission of The Macmillan Company, publishers.

- I. Proprietors
- II. Professional service
- III. Managerial service
- IV. Commercial service
- V. Clerical service
- VI. Agricultural service
- VII. Artisan-proprietors
- VIII. Building and related trades
- IX. Machine and related trades
- X. Printing trades
- XI. Miscellaneous trades in manufacturing and mechanical industries
- XII. Transportation service
- XIII. Public service
- XIV. Personal service
- XV. Miners, lumber-workers, and fishermen
- XVI. Common labor
- XVII. Occupation unknown

Sims in his "Score Card for Socio-Economic Status" uses a fivefold classification of occupations, which differs slightly from the Taussig classification (21, p. 22):

Group I. Professional men, proprietors of large businesses, and higher executives.

Group II. Commercial service, clerical service, large land-owners, managerial service of a lower order than in Group I, and business proprietors employing from five to ten men.

Group III. Artisan proprietors, petty officials, printing trades employees, skilled laborers with some managerial responsibility, shop-owners and business proprietors employing one to five men.

Group IV. Skilled laborers (with exception of printers) who work for some one else, building trades, transportation trades, manufacturing trades involving skilled labor, personal service. Small shop-owners doing their own work.

Group V. Unskilled laborers, common laborers, helpers, "hands," peddlers, varied employment, vendors, unemployed (unless it represents the leisure class or retired).

The attempt to scale occupations according to economic-social level received aid from the results of intelligence testing carried on in the United States Army during the World War. It was found, after the scores on Army Alpha were tabulated by occupations, that the intelligence level of an occupation was a very good index of its social and economic importance. The following table gives occupational intelligence standards as derived from

540 Diagnosing Personality and Conduct

testing with Army Alpha during the war and corrected by Fryer in later research carried on in the Central Branch of the Brooklyn Y. M. C. A.

TABLE 95

OCCUPATIONAL INTELLIGENCE STANDARDS BASED ON ARMY ALPHA INTELLIGENCE TESTS

Score average	Score range	Occupation	Score average	Score range	Occupation
161	110-183	Engineer	69	48-94	Handyman (gen'l. mechanic)
152	124-185	Clergyman			
137	103-155	Accountant	69	46-90	Policeman and Detective
127	107-164	Physician			
122	97-148	Teacher	68	51-97	Auto assembler
119	94-139	Chemist	68	47-89	Engineer (marine)
114	84-139	Draftsman	68	42-86	Riveter (hand)
111	99-163	Y. M. C. A. secretary	67	50-92	Toolmaker
			66	45-92	Auto engine mechanic
110	80-128	Dentist			
109	81-137	Executive (minor)	66	45-91	Laundryman
103	73-124	Stenographer and typist	66	49-86	Gunsmith
			66	44-88	Plumber
101	77-127	Bookkeeper	66	44-88	Pipe-fitter
99	78-126	Nurse	65	44-91	Lathe hand (production)
96	74-121	Clerk (office)			
91	69-115	Clerk (railroad)	65	43-91	Auto mechanic (general)
86	59-107	Photographer			
85	57-110	Telegrapher and radio operator	65	43-91	Auto chauffeur
			65	42-89	Tailor
83	64-106	Conductor (railroad)	65	44-88	Carpenter (bridge)
			64	43-83	Lineman
82	57-108	Musician (band)	63	40-89	Machinist (general)
81	59-106	Artist (sign letterer)	63	46-88	Motorcyclist
81	60-106	Clerk (postal)	63	41-86	Brakeman (railroad)
81	57-109	Electrician	62	31-94	Actor (vaudeville)
80	62-114	Foreman (construction)	61	40-85	Butcher
			61	44-84	Fireman (locomotive)
80	56-105	Clerk (stock)			
78	54-102	Clerk (receiving and shipping)	61	39-82	Blacksmith (general)
			60	38-94	Shop mechanic (railroad)
78	61-106	Druggist			
77	59-107	Foreman (factory)	60	36-93	Printer
75	56-105	Graphotype operator	60	40-84	Carpenter (general)
74	53-91	Engineer (locomotive)	59	40-87	Baker
			59	39-83	Mine drill runner
72	54-99	Farrier	59	38-81	Painter
70	46-95	Telephone operator	58	37-85	Concrete worker
70	44-94	Stock checker	58	40-83	Farmer
69	49-93	Carpenter (ship)	58	37-83	Auto truck chauffeur

<i>Score average</i>	<i>Score range</i>	<i>Occupation</i>	<i>Score average</i>	<i>Score range</i>	<i>Occupation</i>
58	37-82	Bricklayer	48	21-89	Station agent (general)
57	41-81	Caterer			
57	39-71	Horse trainer	40	19-67	Hospital attendant
56	38-76	Cobbler	40	19-60	Mason
55	35-81	Engineer (stationary)	35	18-62	Lumberman
			35	19-57	Shoemaker
55	34-78	Barber	32	16-59	Sailor
55	35-77	Hostler	31	20-62	Structural steel worker
52	38-96	Sales clerk			
52	33-74	Horse-shoer	31	19-60	Canvas maker
51	31-79	Storekeeper (factory)	30	16-41	Leather worker
			27	19-63	Fireman (stationary)
51	26-77	Aëroplane worker	27	17-57	Cook
51	31-74	Boiler-maker	26	18-60	Textile worker
50	33-75	Rigger	22	16-46	Sheet metal worker
50	30-72	Teamster	21	13-47	Laborer (construction)
49	40-71	Miner (general)	20	15-51	Fisherman

Concerning this list Fryer (15) says,

"The mean for the occupation is presented as the 'score average,' and the 'score range,' indicating the range of intelligence within which can be expected success in the occupation, secures its limits (usually, but not always) from the first and third quartiles. The scores are so presented as to indicate that in all probability an individual must have an intelligence rating within the 'score range' for achievement in the occupation, with the further probability that he should be above the 'score average' to be sure of sufficient intellectual capacity for the occupation."

Examination of the table, however, leads to the conviction that even such a detailed analysis does not present sufficient evidence to make it possible to scale occupations at all accurately. One notes, for instance, that the range of the middle 50 per cent of band musicians is from a score of 57 to a score of 108 on Army Alpha. Twenty-five per cent have scores lower than 57, and twenty-five per cent have scores higher than 108. In other words "band musicians" probably covers nearly the whole range of scores on Army Alpha. In order to place a man accurately on the scale, one must know more than his occupation title, and even with an accurate knowledge of his occupation, judgment must be exercised in placing an individual on the scale.

542 Diagnosing Personality and Conduct

Barr tried to overcome this difficulty by building up a true scale of occupations, based primarily on the intelligence standards of occupations. He prepared a "list of one hundred representative occupations, each definitely and concretely described, and had thirty judges rate them on a scale of zero to one hundred according to the grade of intelligence which each was believed to demand." These ratings were then treated statistically and P. E. values assigned to each occupation.

BARR SCALE OF OCCUPATIONAL STATUS *

<i>P. E. value</i>	<i>Occupation</i>	<i>Description</i>
0.00	Hobo	
1.54	Odd jobs	
2.11	Garbage collector	
3.38	Circus roustabout	Does heavy, rough work about circus.
3.44	Hostler	Care of horses in livery, feed, and sales stables.
3.57	Railroad section hand	Replaces ties, etc., under supervision.
3.62	Day laborer	On street, in shop or factory as roustabout.
3.99	Track layer	Does heavy work under supervision.
4.20	Waterworks man	A variety of odd jobs, all unskilled.
4.29	Miner	Digger and shoveler, etc.
4.81	Longshoreman	Loads and unloads cargoes.
4.91	Farm laborer	Unskilled and usually inefficient.
4.98	Laundry worker	Various kinds of work in laundry (practically unskilled).
5.27	Bartender	
5.41	Teamster	
5.44	Sawmill worker	Heavy work, little skill required.
5.59	Dairy hand	Milking, care of stock under supervision.
5.81	Drayman	
5.87	Deliveryman	Delivers groceries, etc., with team or auto.
6.14	Junkman	Collector of junk.
6.42	Switchman	Tending switch in railroad yards.
6.66	Smelter worker	Metal pourers, casting collectors, etc.
6.27	Tire repairer	In general automobile repair shop.
6.85	Cobbler and shoemaker	Repairman in shoe-shop.
6.86	Munition worker	Average.
6.92	Barber	Not owner. Has charge of chair.
6.93	Moving picture operator	Operates machine which projects pictures.
7.02	Vulcanizer	Understands the process of hardening rubber.

* From Terman, L. M., *Genetic Studies of Genius*, Vol. I, *Mental and Physical Traits of a Thousand Gifted Children*, 2d ed. (Stanford University Press, 1925), pp. 67-69.

Measures of the Environment

543

P. E.

<i>value</i>	<i>Occupation</i>	<i>Description</i>
7.05	General repair man	Repairs broken articles. Uses wood-working tools.
7.06	Ship rigger	Installing cordage system on sailing vessels, working under supervision.
7.17	Telephone operator	
7.19	Cook	In restaurant or small hotel.
7.23	Street-car conductor	
7.24	Farm tenants	On small tracts of land.
7.30	Brakeman	On freight or passenger trains.
7.33	City fire-fighter	Handles the ordinary fire-fighting apparatus.
7.39	Railroad fireman	On freight or passenger train.
7.54	Policeman	Average patrolman.
7.71	Structural steel worker	Heavy work demanding some skill.
7.73	Telephone and telegraph lineman	
7.77	Bricklayer	
7.79	Butcher	Not shop-owner. Able to make cuts properly.
7.91	Baker	
8.02	Metal finisher	Polishes and lacquers metal fixtures, etc.
8.04	Plasterer	Knowledge of materials used necessary.
8.08	General painter	Paints houses, buildings, and various structures.
8.22	Harness maker	
8.40	Tinsmith	Makes vessels, utensils, etc., from plated sheet metal.
8.49	Letter carrier	
8.50	Forest ranger	
8.58	Stone mason	
8.75	Plumber	Average trained plumber employee.
8.89	Gardening, truck farming	Owns and operates small plots.
8.99	Electric repair man	Repairs electric utensils, devices, and machines.
9.28	Bookbinder	Sets up and binds books of all sorts.
9.37	Carpenter	Knows wood-working tools. Can follow directions in various processes of wood construction work.
9.37	Potter	Makes jars, jugs, crockery, earthenware, etc.
9.54	Tailor	Employee in tailoring shop.
9.72	Salesman	In dry-goods, hardware, grocery stores, etc.
10.11	Telegraph operator	In small town.
10.21	Undertaker	In small town. Six months to a year of especial schooling.
10.26	Station agent	In small town. Acts as baggage man, freight agent, operator, etc.
10.26	Mechanical repair man	In shop or factory. Keeps machines in condition.
10.29	Dairy owner and manager	Small dairy, 50-100 cows.

544 Diagnosing Personality and Conduct

<i>P. E.</i>	<i>Occupation</i>	<i>Description</i>
10.53	Metal pattern-maker	
10.54	Wood pattern-maker	
10.54	Lithographer	Makes prints from designs which he puts on stone.
10.76	Linotype operator	
10.83	Photographer	City 1,000-5,000. A few months' training, experience in studio.
10.86	Detective	Traces clues, etc. Employee of detective bureau.
10.99	Electrotyper	
11.17	Traveling salesman	Sells drugs, groceries, hardware, dry-goods, etc.
11.34	Clerical work	Bookkeepers, recorders, abstractors, etc.
11.35	Railroad passenger conductor...	
11.51	Storekeeper and owner	Small town retail dealer, general or special store.
11.74	Foreman	Small factory, shop, etc.
11.78	Stenographer	Writes shorthand and uses typewriter.
12.02	Librarian	In small institution or public library.
12.06	Nurse and masseur	Graduate.
12.74	Chef	Employed in large first-class hotels.
12.84	Editor	Small paper, considerable job work.
12.89	Primary teacher	No college training, two years' special training.
12.96	Landscape gardener	
13.08	Grammar grade teacher	Normal graduate, expects to make profession teaching.
13.20	Osteopath	Training equal to college graduate.
13.21	Pharmacist	In town of from 1,000-5,000 population.
13.29	Master mechanic	Thorough knowledge in his field of mechanics.
13.30	Music teacher	2-4 years' special training, not college graduate.
13.31	Manufacturer	Employs from 10-50 men. Makes simple articles.
13.54	Dentist	Graduate. Two to five years' experience in small town.
13.58	Art teacher	In high school. Three or four years' special training.
13.71	Surveyor	Transit man. City or county surveyor.
13.31	Train dispatcher	Must be mentally alert.
14.45	Land-owner and operator	Very large farms or ranches.
14.70	Musician	Successful player or singer in good company.
15.05	Secretarial work	Private secretary to high state or national officials.
15.14	High school teacher	College or normal graduate. Not the most progressive.
15.15	Preacher	Minister in town of 1,000-5,000. College graduate.

P. E.
value Occupation
Description

15.42	Industrial chemist	Thorough knowledge of the chemistry of manufacturing processes.
15.43	Mechanical engineer	Designs and constructs machines and machine tools.
15.71	Teacher in college	Degree A.B. or A.M. Not the most progressive.
15.75	Lawyer	In town of moderate size. Income \$1,000-\$5,000.
15.86	Technical engineer	Thorough knowledge of the processes of an industry.
16.18	Artist	High-class painter of portraits, etc.
16.26	Mining engineer	Thorough knowledge of mining and extraction of metals.
16.28	Architect	Training equal to college graduate.
16.58	Great wholesale merchant	Business covering one or more States.
16.59	Consulting engineer	In charge of corps of engineers.
16.64	Educational administrator	Superintendent in city of 2,000-5,000. College or normal graduate.
16.71	Physician	Six to eight years' preparation above high school. Income \$5,000 and up.
16.91	Journalist	High-class writer or editor.
17.50	Publisher	High-class magazine and newspaper or periodical, etc.
17.81	University professor	Has A.M. or Ph.D., writes, teaches, and does research.
18.06	Great merchant	Owns and operates a million-dollar business.
18.14	Musician	(Paderewski).
18.33	High national official	Cabinet officers, foreign ministers, etc.
18.85	Writer	(Van Dyke).
19.45	Research leader	Like Binet or Pasteur.
19.73	Surgeon	(Mayo Brothers).
20.71	Inventive genius	(Edison type).

The Barr scale, used by Terman in his study of gifted children, apparently provides the truest means for scaling occupations that we now have. The scale by definition places each occupation in order of the intelligence it demands. Although the matter has not been investigated, it is presumed that such a scale will correlate very highly with a scale made strictly to fit socio-economic level.

In the Minnesota Mechanical Ability Project a scale similar in nature to the Barr scale was drawn up to be used in grading occupations according to the amount of mechanical ability involved.

Measures of the Home

There are two fundamental methods of obtaining data necessary to measure the social-economic level of the home. One is to have a visitor enter the home of a given child for the purpose of noting on a prepared schedule the equipment and living conditions which he finds. The other is to ask the child questions designed to yield similar information. The former method, although more expensive in time and energy, yields the more reliable results. The latter method can be used in large groups in school and the information can be more quickly and easily obtained, but it is liable to inaccuracy.

As frequently occurs, one of the earliest pieces of work in the technique of measuring the home was one of the best, though it has since fallen into oblivion. Perry (20) in 1912, wishing to study the standard of living, sought a measure of the adequacy of the home. A preliminary study convinced him that the four rooms of a house in descending order of importance were the kitchen, bedroom, dining-room, and parlor. His plan proposed to inventory the furnishings of these four rooms with a view toward determining how adequately they served the needs of the household. He provided a check list for the furnishings of each of these four rooms, adding a system of weights to aid in assigning ratings for each article checked. The sum of the ratings may be used as a measure of the level of the home. The investigator merely passes through the house, checking off on the schedule the articles in evidence. The ratings may be computed at leisure. The score is the sum of the weightings assigned, divided by the sum of the required weights, which represents a sort of minimum standard for an adequate household.

A similar method has been employed by Chapin (7) in measuring the equipment of the living-room of an urban middle-class family. His scale consisting of fifty-three items is divided into four sections:

- I. Fixed features
- II. Built-in features
- III. Standard furniture
- IV. Furnishings and cultural resources

Each item either contains directions for assigning credit or includes a brief scale to enable the visitor to give it its proper weight.

Examples of items are:

2. Floor covering.
Composition 1, carpet 2, small rugs 3, large rugs 4.
21. Chair
Straight, rocker, arm-chair, high-chair, 1 each.
49. Radio
Crystal 1, one-tube 2, two-tube 3, three-tube 4, super-heterodyne 5, etc.

Chapin chose the living-room because it could be scored in a comparatively short interview which would not involve objectionable inquisitorial questions. He found that the score of the living-room had a high correlation with the general socio-economic status of the family. Indeed, this measure of the living-room has as high a correlation with an average of several measures of socio-economic level as any one of the scales specially designed to measure general socio-economic level has with the average.

Commons (11), still earlier, developed a score card for measuring the home. This score card, in two sections, is so arranged that the maximum credit possible for each section is 100 points. Considerable judgment must be exercised in using this type of score card for items in which less than the maximum credit is granted. Until one has had much experience with the scale and has built for himself a set of personal standards, the assignment of credit is extremely subjective. Commons tries to guide the assignment of credit by the following key:

Instructions for Discrediting When Depending on Judgment

Deduct from possible 6; very slight 1; slight 2; marked 3; very marked 4; extreme 5.

Deduct from possible 3; very slight $\frac{1}{2}$; slight 1; marked $1\frac{1}{2}$; very marked 2; extreme $2\frac{1}{2}$.

A brief outline of the score card follows:

Dwelling House Score Card

I. Dwelling	
Location	18
Congestion of buildings	26
Window openings	11
Air and ventilation	13
Structural conditions	6
House appurtenances	26
	<hr/>
	100

II. Occupants

Congestion of occupancy	61
Condition of air and ventilation	18
Cleanliness	21
	<hr/>
	100

A third method of measuring the home is represented by the "Whittier Scale for Grading Home Conditions" proposed by Williams (29). This is a series of five five-point rating scales for measuring the factors of *necessities*, *neatness*, *size*, *parental conditions*, and *parental supervision*.

This scale has the problem of conduct directly in mind, since it contains not only items relating to the adequacy and furnishing of the house, but two items pertaining to the social adequacy of the home and the parental control. We note in reading the items on the scale that they apply to a rather specific situation. Workers who wish to use this type of scale should redefine the items to fit the locality in which they intend to work.

Measures of Socio-economic Level

Early work by Van Denberg (26), Holley (17), Counts (12), and Kornhauser (18) paved the way for the later work of Chapman and Sims (8). Chapman, of Yale, was attracted to the possibility and value of studying intensively the significance of various factors in the environment of a child which could best serve as indices of this more general thing which we call socio-economic level. Where previous workers had used such factors as parental occupation, telephone in the home, etc., Chapman conceived the idea of trying out a large number of items, correlating each of the items with the most valid criterion of socio-economic level obtainable, and selecting the most significant items to form a scale for the measurement of socio-economic level.

Chapman, in collaboration with Sims, one of his students, undertook to study the factors which influence participation in extra-curricular activities, using a questionnaire of sixteen questions relating to socio-economic status. This questionnaire was given to the students (over 3,000) enrolled in a New Haven high school. Correlations were found between each item (using biserial r , since questions were answered *yes* or *no*) and a total

WHITTIER SCALE FOR GRADING HOME CONDITIONS (29, p. 283) (Whittier State School, Department of Research)

NECESSITIES	5	NEATNESS	5	SIZE	5	PARENTAL CONDITIONS	5	PARENTAL SUPERVISION	5
Plenty of high-grade furniture. Home well equipped. Piano, phonograph, good pictures, modern improvements. 7-passenger touring car. Father a real-estate dealer, good income.		Clean, well arranged. Everything neat and tidy. Well-kept lawn; nothing scattered about. House well ventilated.		4 persons, occupy two-story house, with 8 or 9 rooms, all large. Boy has room of his own.		Parents living together, and in perfect harmony. Have not been separated.		Parents at home evenings, where boy is kept, and supplied with good books and wholesome surroundings. Parents take much intelligent interest in his welfare.	
NECESSITIES	3	NEATNESS	3	SIZE	3	PARENTAL CONDITIONS	3	PARENTAL SUPERVISION	3
Sufficient furniture, of a practical sort, but plain, not expensive. Pianola. Income, proceeds from rooming house and day labor of father.		Fairly well arranged. Clean, but not with special caution. Wall not papered, but not dirty.		2 persons, occupy three-room cottage. Living-room, dining-room, and kitchen. Disappearing bed in dining room and couch (for boy) in living-room.		Parents separated. Mother remarried, and has her boys with her.		Father away all day, at work. Mother keeps house. Has good discipline in the home, but it does not extend beyond.	
NECESSITIES	1	NEATNESS	1	SIZE	1	PARENTAL CONDITIONS	1	PARENTAL SUPERVISION	1
Bare essentials. Plain wooden kitchen chairs, no rockers, light table, no pictures or decorations except advertising calendar. Income, county support only.		Dirty floors. Plaster falling from walls and ceiling. Windows dirty, dust over everything. Children dirty.		5 persons, occupy three-room cottage. Family eats in kitchen. One bedroom, screened off kitchen. Living room also used for sleeping purposes.		Mother dead, father remarried to a girl four years older than the boy. Father and mother also holic.		Parents lazy, shiftless, ugly toward the boy. Taught him to steal. Have no interest in his development.	

550 Diagnosing Personality and Conduct

score found by using all the items and weighting each equally. From these data a crude scale was developed.

After the death of Chapman, Sims (21) continued the work on a larger scale, his work being financed by the Character Education Inquiry. Sims used the following criteria in assembling questions for his original schedule:

1. Each question must be indicative of the economic or the cultural level of the home or both.
2. The questions must cover as many aspects of the home background as possible.
3. The questions must be stated so that the child can understand them.
4. The questions must ask for information which the child is willing to furnish.
5. The questions must ask for information which the child can furnish.
6. The questions must be stated in such a manner that there will be a minimum chance of error.
7. The questions must allow of answers that are comparable.

Fifty-six questions were used in the original try-out. The items which gave the highest bi-serial correlations with the composite are as follows:

TABLE 96

BI-SERIAL CORRELATION COEFFICIENTS BETWEEN ASPECTS OF HOME BACKGROUND AND SOCIO-ECONOMIC STATUS OF THE HOME

(Sims, p. 12)

1. Servants865	9. Books773
2. Golf860	10. Father h. s.764
3. Father's occupation856	11. Lectures758
4. Dancing lessons811	12. Mother goes to lectures....	.743
5. Dental work793	13. Rooms people739
6. Music concerts788	14. Mother college735
7. Bank account786	15. Furnace731
8. Mother h. s.775	16. Vacation705
17. Telephone.....	.702		

However, some of these items were answered *yes* by such a small percentage of a normal group as to be less useful than other items. Sims used the following criteria in selecting items for his scale (21, p. 18):

1. The ability and willingness of the persons tested to furnish the information.
2. The correlation of a given item with the total of the other

questions. Other things being equal, the higher the correlation, the more desirable the question.

3. The intercorrelation between the items. Other things being equal, the lower the intercorrelation the more desirable the question.

4. The percentage of the population possessing the article or engaged in the activity asked about in the question. Other things being equal, it is desirable to have the questionnaire include as wide a representation of items as possible; that is, include some items possessed by many, others by few; otherwise there would be a tendency for a large undistributed group to form at one or the other extreme.

5. The reliability of the question.

6. The variety of aspects of home background recommended. Other things being equal, it is desirable to measure as many aspects of the complex as is possible. Where there are two or three questions pertaining to the same aspect, for example, of two questions asking for information as to lectures, it is desirable to retain only the better one.

7. Common sense.

Twenty-seven items were finally selected for the questionnaire. The scoring method is rather elaborate, being based essentially on a method devised by Chapman.

1. The percentage of those answering a question *yes* and those answering *no* was obtained.

Do you have a telephone in the house?	Yes	No
	38.5%	61.5%

2. Using the Kelley-Wood table of the probability integral, these percentages are transferred into sigma values:

$$+0.993 \quad -0.621$$

3. These sigma values are multiplied by certain weights that have been assigned to each question on the basis of correlation with the criterion intercorrelations, reliability, etc. The weighting for the question about the telephones is 5. Multiplying the sigma values by 5, one has:

$$+4.965 \quad -3.105$$

4. Add 10 to each item to make all values positive:

$$+14.965 \quad +6.895$$

5. Give approximate integral value:

$$+15 \quad 7$$

These numbers (15 and 7) are used as scores to be given to *yes* and *no* answers to this question.

Sims reports a reliability of $+.94$, correlating the responses of 200 paired siblings.

The care, thoroughness, and soundness of method used in constructing this scale make it the most valid method available at present for determining socio-economic level. In his monograph, Sims reports the findings in different communities which permit accurate comparisons between them.

Testing for Cultural Background

A novel method of determining the adequacy of the home environment, developed in connection with the work of the Character Education Inquiry, consists in testing the cultural responses of children. For this purpose there was developed a "Good-Manners Test" by Orr (16) and an "Apperception Test" by Burdick (4). The former is a test of manners and etiquette employing true-false, yes-no, and best-answer techniques. The latter, employing a variety of techniques, endeavors to tap the richness of the child's home life by questioning him in an indirect way about its economic and cultural resources. Reliability coefficients of .621, .708, .741, and .766 were found by correlating two scores of the test, and a reliability of .499 by correlating the scores of siblings. The test correlates with a criterion of home background .659.

These correlations indicate that this approach has promising possibilities. An inspection of the tests themselves, however, indicates that they probably correlate well with intelligence, and we are left wondering to what degree their value as a measure of home background is reached through the intermediary of intelligence. There is another still more fundamental objection to this type of approach to the measurement of the environment in that it confuses the environment with the verbal responses which the environment stimulates. In a sense, it is begging the question to assume that one can be used as a measure of the other. Indeed, it is extremely important to know exactly how much relationship there is between the environment and a child's responses to it. We feel safer, therefore, in studying the environment directly rather than through its effects.

Among other attempts to measure special phases of home background should be mentioned the "Questionnaire on Cultural and Mechanical Environment" which was developed and used as a part of the "Minnesota Mechanical Ability Study" (19). One section of this questionnaire deals with the "tools in the home,"—a list of tools to be checked for ownership by father and by son, yielding an inventory and measure of the mechanical resources of the home. This check list possesses high reliability (.94 and .89).

REFERENCES

1. BARR, F. E., *A Scale for Measuring Mental Ability in Vocations and Some of Its Applications* (Stanford University Press, 1918).
2. ———, "Barr Social Rating Scale in Occupational Status," in Terman, L. M., *Genetic Studies of Genius*, Vol. I, 2d Ed. (Stanford University Press, 1925).
3. BRIDGES, V. W., and COLER, E., "The Relation of Intelligence to Social Status," *Psychological Review*, 24: 1-31 (1917).
4. BURDICK, E. M., *A Group Test of Home Environment*, Archives of Psychology, No. 101 (Oct., 1928).
5. CHAPIN, F. S., "A Quantitative Scale for Rating the Home and Social Environment of Middle Class Families in an Urban Community," *Journal of Educational Psychology*, 19: 99-111 (Feb., 1928).
6. ———, *Field Work and Social Research* (The Century Company, 1920), ch. VII.
7. ———, "Measuring the Volume of Social Stimuli: A Study in Social Psychology," *Social Forces*, 1: 479-495 (March, 1926).
8. CHAPMAN, J. C., and SIMS, V. M., "The Quantitative Measurement of Certain Aspects of Socio-economic Status," *Journal of Educational Psychology*, 16: 380-390 (1925).
9. CLARK, W. W., "Scale for Grading Social Conditions," *Journal of Applied Sociology*, 7: 13-18 (1922-1923).
10. ———, "Whittier Scale for Grading Juvenile Offenses," Whittier, Calif., Bureau of Juvenile Research Bulletin, No. 11 (1922).
11. COMMONS, J. R., "Standardizing the Home," *Journal of Home Economics*, 2: 23-29 (Feb., 1910); also "Standardization of Housing Investigations," in *Quarterly Publications of the American Statistical Association*, 11: 319-326 (Dec., 1908).
12. COUNTS, G. S., *The Selective Character of American Secondary Education*, Supplementary Educational Monographs, No. 19 (University of Chicago Press, 1922).

554 Diagnosing Personality and Conduct

13. COUNTS, G. S., "The Selective Principle in American Secondary Education," I, II, *School Review*, 29: 657-667 (1921); 30: 95-109 (1922).
14. FLEMMING, C. W., and RUTLEDGE, S. A., "The Importance of the Social and Economic Quality of the Home for Pupil Guidance," *Teachers College Record*, 29: 202-215 (Dec., 1927).
15. FRYER, D., "Occupational—Intelligence Standards," *School and Society*, 16: 273-276 (Sept. 2, 1922).
16. HARTSHORNE, H., and MAY, M. A., *Studies in Deceit—Studies in the Nature of Character*, I (The Macmillan Company, 1928).
17. HOLLEY, C. E., "The Relationships between Persistence in School and Home Conditions," *15th Yearbook of the National Society for the Study of Education* (1916), pt. 2, 119 pp.
18. KORNHAUSER, A. W., "The Economic Standing of Parents and Intelligence of Their Children," *Journal of Educational Psychology*, 9: 159-164 (1918).
19. PATERSON, D. G., ELLIOTT, R. M., and others, *Minnesota Mechanical Ability Tests* (The University of Minnesota Press, 1930).
20. PERRY, C. A., "A Measure of the Manner of Living," *Journal of the American Statistical Association*, 13: 398-403 (1912-1913).
21. SIMS, V. M., *The Measurement of Socio-Economic Status* (Public School Publishing Company, 1928).
22. SLAWSON, J., *The Delinquent Boy—A Socio-psychological Study* (Richard G. Badger, The Gorham Press, 1926).
23. SYDENSTRICKER, E., and KING, W. I., "The Measurement of the Relative Economic Status of Families," *Journal of the American Statistical Association*, 17: 842-859 (Sept., 1921).
24. TAUSSIG, F. W., *Principles of Economics*, 1st ed. (The Macmillan Company, 1911), II, 134-137.
25. United States Bureau of the Census, *Index of Occupations—Alphabetical and Classified* (1920).
26. VAN DENBURG, J. K., *Causes of the Elimination of Students in Public Secondary Schools of New York City*. Teachers' College Contributions to Education, No. 47 (1911).
27. WAPLES, D., "Indexing the Qualifications of Different Social Groups for an Academic Curriculum," *School Review*, 32: 537-546 (1924).
28. WILLIAMS, J. H., *A Guide to the Grading of Homes*, Whittier, Calif., Bureau of Juvenile Research Bulletins, No. 7 (1918).
29. ———, "The Whittier Scale for Grading Home Conditions," *Journal of Delinquency*, 1: 273-286 (1906).
30. WYLIE, A. T., in Farrand, W., and O'Shea, M. V., *The All-Year Schools of Newark, N. J.*, Part IV (1925).

Chapter XVI

THE CASE STUDY: A COMPREHENSIVE STUDY OF THE INDIVIDUAL

WITH the development of social work, and with the increased attention being given to individuals in education, more and more consideration is being devoted to case work and the techniques of the case study. The word *case* in this connection has several ill-defined meanings. The *case method*, long used in law schools, has percolated through other types of professional training and has even penetrated to college and secondary school instruction. It is a method of instruction employing the study of cases as the source of problem material and as a means of making concrete the application of general principles. *Case work* refers to remedial, therapeutic, or corrective work carried on by physicians or social workers in efforts to bring about better adjustments among individuals. This latter use of the word *case* is closely linked up with the *case study*, by which is meant a comprehensive, exhaustive investigation of an individual. The case study and case work usually are closely intermingled. The case worker may give advice and propose remedies even while she is still proceeding with the collection of pertinent facts. In the present discussion, however, the *case study* alone is considered.

It should be emphasized at the outset that the *case study* is not a research method. Primarily its function is to study the individual with a view toward helping him. If the case study yields evidence that is helpful in scientific investigation, this is only a by-product and not its main contribution. If the case study employs a schedule of facts to be noted, and if these facts have been obtained in a reliable, objective manner, then these data may be used in research investigations. But the case study method contains no guarantee that its observations are complete or uniform, or that scientifically valid methods were employed in getting them. The case study has the individual's interest uppermost

556 Diagnosing Personality and Conduct

in mind and may or may not employ a regular inquiry schedule or use consistent methods.

The case study is just as good as and no better than the methods employed in gathering the data. To most persons the case study means the interview. If the interview method of obtaining data for a case study is used, then the case study possesses the validity and reliability of the interview method. However, the case study does not need to employ the interview method solely, and when objective tests or questionnaires are available they should be used in preference to the more subjective interview. Again the case study usually means the study of one individual at a time, but if group methods yield as reliable evidence, there is no reason why group tests or questionnaires should not be used as part of the data for case studies. The case study ought to employ the most objective, the most reliable, the most economical, in short, the best methods of collecting data available. When this is done, the data obtained in making a case study are welcomed as data to be used in scientific research.

To be more concrete, medical science is now recommending the periodic health examination, which is a case study of the health and physical condition of an individual. As part of this examination the subject may be asked questions about his feelings or his living habits. But the main part of the examination consists of an objective personal examination of the individual's physical condition, using the most refined techniques and measuring devices that medical science has produced. The examinations are thorough and the conclusions and recommendations are based on a survey of all the findings. That is a case study. As the records accumulate they become invaluable data for scientific research into the incidence and etiology of disease. In the diagnosis of personality and conduct similar standards of objectivity and reliability should be adopted.

In planning a case study, careful consideration must be given to the items to be included in the schedule. If the aim is a complete psychological case study, the practice is usually to include inquiry into every factor that might have significance in understanding the case. In such a comprehensive schedule questions will be included relating to heredity, physical constitution, developmental history, intelligence and other special abilities, home

and other environmental factors, favorite activities, companions, personal adjustments, behavior characteristics, emotional stability, reputation as reported by parents, teachers, and friends, ideals, ambitions, wishes, and possibly an intensive study of emotional complexes. Many of these branches of the inquiry will yield nothing of particular significance for an understanding of the peculiarities of the case, but all must be investigated in the interests of thoroughness. Quite possibly the clue to the problem will be found to be in any one of the categories listed above, and probably the total syndrome will reveal abnormal conditions in several. A problem child may reveal bad inheritance, constitutional weaknesses, poor home conditions, all in harmony with evidence as to behavior peculiarities and his own evidence as to maladjustment. In the case study we can afford to neglect nothing.

Usually, however, since one's purpose is the understanding of some particular set of symptoms, it is possible to abridge the exhaustiveness of the study. As scientific inquiry proceeds, evidence accumulates as to the relationship of the various factors going to make a total personality picture. With such evidence at hand it becomes possible to begin an investigation at once with the factors having the greatest probability of relationship. This, of course, is the practitioner's method. The experienced diagnostician, or the student in command of the pertinent scientific evidence bearing on a condition, does not need to undertake an exhaustive and complete case study. He naturally uses certain tests first, and if they show abnormal conditions, he may proceed no further, confident in his assumption that he has the key to the situation.

As an example, consider the diagnosis of reading disability. The symptoms—difficulty with reading—are comparatively limited and distinct. The case study, then, need not be exhaustive, but may proceed at once to the particular functions involved. Tests of deficiencies in vocabulary, word recognition, phrase and sentence comprehension, paragraph comprehension, visual, auditory, and motor functions; associative learning, or even nervous and emotional stability may be employed.* Scientific work has already demonstrated relationship between deficiency in these various functions and difficulties in reading.

* See Gates, A. I., *The Improvement of Reading* (The Macmillan Company, 1928).

In considering what abridgment is possible in the case study, three cautions must be kept in mind. In the first place an investigator must draw conclusions on the basis of *proved* relationships. It is a natural human tendency to generalize from a few cases, particularly the cases with which one has had personal experience. Witness how only too frequently practitioners act on hunches, and play up fads and their pet theories. It is not too much to ask that judgment be based only on relationships that have been demonstrated to exist instead of on merely plausible speculations. Nervousness among children has been variously ascribed to heredity, to undernourishment, to fatigue, to unhappy home conditions, to a sense of inferiority or guilt, and to masturbation. The diagnostician who rides a hobby will investigate his abnormality with respect to one of these factors, and if he finds it, will cease further search.

This leads us naturally to the second caution, which is that a case study should not be abridged when one suspicious factor has been found, particularly when the relationship between this factor and the disturbing symptom is loose and uncertain. Researches do not often yield categorical evidence that such and such a factor is a cause and another one not a cause. Rather they show that it is a matter of degree, that the relationship between factors is low or high, loose or close. In practice, error creeps in through discovering the presence of adverse conditions with respect to a factor that has low relationship to the disturbing symptom and resting there, satisfied that the cause has been located. Where relationships are low, the search should be continued for all possible disturbing factors.

As an example, Hartshorne and May report that negro school-children cheat more on school tests than would be expected from comparisons with other racial groups. This might be falsely generalized into, "All negro children are dishonest." Then a school principal, having in mind this imperfection generalization, might easily reason concerning a pupil sent to him for cheating after this fashion: "This pupil cheated; but he is colored; what can you expect?" The absurdity of this judgment is patent, and yet psychiatrists are daily making case studies, drawing inferences, and making recommendations on evidence that has less foundation than that used in this illustration.

The third caution lies in the other direction. The author's re-

A Comprehensive Study of the Individual 559

view of diagnostic techniques convinces him of the *value of small correlations*. In the early stages of educational research, students were looking for the big, obvious relationships between phenomena, and relationships represented by coefficients of correlation as low as .10 or .20 were considered negligible, unworthy of attention. We may confidently assert that practically all of the marked relationships in human behavior have now been explored. As one example, the relation between intelligence and a myriad of other factors has been investigated. The general conclusion is that the relationship is often close, but always with enough leeway so that other factors may have considerable influence in determining actual behavior or the status of the individual.

There is a need now for the patient investigation of relationships of low degree carried out on sufficiently large numbers and with accurate enough measures to insure that the relationships are reliably determined. These low relationships are now sensed to be of particular importance in diagnosing conduct. It is well known that *extreme values* of factors which have low regression weights in the regression equation are potent for determining an issue. Usually, because of the interrelation between factors, an extreme score in one of the variables tends to pull other factors along with it. As an example, take the relationship between cheating in school and low economic status of the family. If a child comes from a very impoverished family, that fact is apt to carry with it low family cultural standards, and possibly unfavorable race, religion, companions, and whatnot, all tending to work in the direction of habits of dishonesty. Many factors are contributory, and no one factor can properly be isolated as the cause, unless poverty, being the most extreme, can be thought of as dragging the others along with it.

In a similar manner it is said that the common cold is caused by sitting in a draft, though many people sit in drafts without getting colds. Along with the exposure must go a state of body chemistry, the prevalence of colds in the community, and other contributory factors. One factor alone stands out as the inciting cause because it is more extreme than the others. Factors with low relationships become important because of their influence when extreme. And though research needs now to give more careful attention to low relationships, this is not to be taken to mean that the clinical worker is given freedom to ride his hobby.

562 Diagnosing Personality and Conduct

of testing in which sixty-four tests of behavior were administered to 850 school-children; they were able to approach the problem of the measurement of character empirically. They used three criteria of character: (a) reputations of the pupils obtained from both teachers and pupils by a variety of rating devices; (b) a series of 100 character portraits graded by competent persons on the basis of descriptions afforded by testing and ratings; and (c) measures of the consistency of behavior in terms of the smallness of variability of behavior.

With each of these measures there were correlated other tests or indices used in the Character Education Inquiry. Tables 97, 98, and 99 give the main facts as presented in *Studies in the Organization of Character*:

TABLE 97

RELATIVE SIGNIFICANCE OF PERFORMANCE FACTORS FOR TOTAL REPUTATION*
(from Hartshorne, May, and Shuttleworth)

<i>r</i> 's for population Y, by class-room ¹	<i>r</i> 's for population XYZ ²
1. Teachers' marks412	1. Teachers' marks78
2. Opinion A260	2. Opinion B53
3. Opinion B254	3. Opinion A49
4. Good citizenship188	4. Information49
5. Information174	5. Good citizenship40
6. Culture237	6. Culture53
7. Kits + envelopes (service) .283	7. School honesty81
8. Athletic contest (honesty). .240	8. Total honesty48
9. School honesty183	9. Service63
10. Inhibition total210	10. Inhibition total62
11. Persistence total117	11. Persistence total50
12. Emotional stability383	12. Emotional stability38
13. Intelligence (IQ)244	13. Intelligence (MA)35
14. Resistance to suggestion152	14. Resistance to suggestion .. .32
15. Self-function071	15. Self-functioning31
16. Age121	16. Age —.23

¹ Corrected for class-room heterogeneity. ² Corrected for attenuation.

Most significant for the problem of the measurement of character was the finding that *general integration* correlates with the average of twenty-three tests .440, which corrected for attenuation becomes .73; correlates with total reputation .400, which when corrected becomes .79; and correlates with the character portraits .344 which when corrected becomes .55. In other words,

* From Hartshorne, H., May, M. A., and Shuttleworth, F. K., *Studies in the Organization of Character*. By permission of The Macmillan Company, publishers.

A Comprehensive Study of the Individual 563

TABLE 98

RELATIVE SIGNIFICANCE FOR CHARACTER OF MISCELLANEOUS CONCOMITANTS*
(from Hartshorne, May, and Shuttleworth)

Factor	Test	r
Group I Reputation	Conduct record	.716
	Check list	.664
	Scholastic marks	.619
	"Guess who"	.585
	Department marks	.514
Group 2 Moral knowledge and opinion	Total reputation	.611
	Opinion A + B	.602
	Opinion B	.569
	Information	.434
	Opinion A	.424
Group 3 Conduct	Good citizenship	.399
	Service total	.584
	School honesty	.490
	Inhibition (3 tests)	.449
Group 4 Culture	Persistence total	.442
	Burdick	.425
	Intelligence	.391
	Resistance to suggestion	.323
Group 5 Personal factors	Mental age	.317
	Emotional stability	.315
	Self-functioning	.314
	Service tests	
	Kits + Envelopes	.560
	Kits	.375
	Money vote	.274
	Coöperation with class	.129
	Free choice	.095
	Honesty tests	
	IER (answer sheets)	.415
	Coördination (peeping)	.415
Group 6 Separate conduct tests	Athletic contest	.345
	Speed (adding answers)	.276
	Inhibition	
	Total of 4 tests	.384
	Picture inhibition	.142
	Persistence	
	Cross and square	.282
	Persistence for self and class	.239
Group 7 Socio-economic status	Story resistance	.022
	Sims' score card	.140
Group 8 Age	Chronological age	-.074

* From Hartshorne, H., May, M. A., and Shuttleworth, F. K., *Studies in the Organization of Character*. By permission of The Macmillan Company, publishers.

564 Diagnosing Personality and Conduct

TABLE 99

CORRELATION OF GENERAL INTEGRATION WITH MISCELLANEOUS MEASURES*
(from Hartshorne, May, and Shuttlesworth)

CONDUCT			KNOWLEDGE		
	Raw r	Cor- rected r ²		Raw r	Cor- rected r ²
School honesty354	.61	Good citizenship.	.504	.87
Service total179	.38	Information315	.53
Inhibition total ¹296	.74	Opinion A B340	.58
Persistence total..	-.066	-.13			
REPUTATION			ABILITY AND STATUS		
	Raw r	Cor- rected r ²		Raw r	Cor- rected r ²
Teachers' marks...	.361	.70	CAVI intelligence		
Department232		(sigma)120	.20
Conduct record...	.352		Resistance to sug- gestion246	.49
'Guess who'223		Emotion stability	.194	.37
Total reputation..	.400		Self-functioning .	.289	.51
			Age	-.041	-.06
			Sims (socio-eco- nomic)138	.24
			Burdick (culture)	.184	.33

¹Omitting the Picture inhibition test.

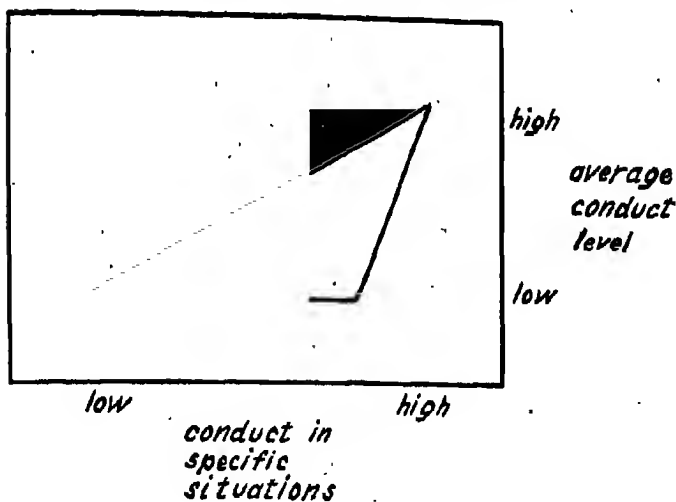
²Corrected for attenuation.

the more desirable the qualities a person has in general, the more consistent he is; and the more undesirable the qualities a person has in general, the more inconsistent he is. The man who is generally honest is also consistently honest—that is to say, he is dependably honest. But the man who is generally dishonest is inconsistent—he may be honest in one situation and dishonest in another. The man with good character seems to be organized from within; at least, he acts uniformly and consistently with regard to commendable things. But the man with low character is more a creature of chance and impulse and is blown about by every change of circumstance.

*From Hartshorne, H., May, M. A., and Shuttlesworth, F. K., *Studies in the Organization of Character*. By permission of the Macmillan Company, publishers.

A Comprehensive Study of the Individual 565

The upshot of this finding is that a measure of the character of a person whose character is high is valuable, for since the person's conduct is *consistent*, one can tell what he will do relatively accurately in specific situations. A knowledge that a man's character is high would be of great importance in employing a bank teller, for it would be as good a guarantee as could be obtained that the person is *dependable* and tends to be more consistently honest, conscientious, thorough, accurate, and the



like in most situations in which he finds himself than the average man. On the other hand, knowledge that a man's character is low tells us more than anything else that he is not dependable and that one can predict little about his conduct in specific situations. He might be honest, and he might not. Probably he would tend to adopt the group mores, but would yield somewhat readily to temptation.

In summarizing Hartshorne, May, and Shuttleworth's results, it appears that the two methods best adapted for estimating a person's conduct are *ratings* and *tests of conduct knowledge and opinion*. Performance tests would also be useful if they were less awkward and expensive, both in time and in money.

That ratings and paper-and-pencil tests of knowledge and opinion should finally emerge as the most significant indices of character may seem something of an anomaly. After the World

War, rating methods were so harshly criticized that it was widely proclaimed that they were forever discredited. However unreliable they may be, and however much they may be biased by halos of general impression, they emerge as one of the best methods of character diagnosis. To be sure, since the war there have been several improvements in rating techniques, the most important of which were developed by the Character Education Inquiry itself. It is significant that ordinary school marks operate as one of the most effective single measures of character.

It may also strike the casual observer as strange that tests of knowledge and opinion should finally precipitate as a superior measure of character when it has been demonstrated that there is so little relationship between knowledge of conduct and conduct itself. However, we must remember that character is only an average and that the one thing that perhaps helps most to determine the average level of conduct and also the integration of character, is one's verbal organization. The evidence strongly supports the conclusion that character is integrated largely through verbal organization. The person who knows what is best and right to do tends on the whole to be the one who does the best thing and the right thing. He who does not know what is best to do is impelled more by force of circumstances, the *mores*, and by habit. To know that a man scores high on a moral knowledge test usually tells much concerning his character and conduct; to know that a man scores low tells us little.

One must agree with Hartshorne, May, and Shuttleworth that to get the best estimate of character, a battery or composite of tests should be used. These workers propose two such possible batteries, one a group of ten tests including the "guess who," check list, teachers' marks, deportment, opinion A and opinion B, kits and pictures, coördination, inhibition (three tests), persistence total, and Burdick culture. This group of ten tests correlates .750 with their "character portraits." A larger battery of twenty-three measures correlated .721 with the criterion, which was raised to .81 by correction for attenuation. However, the composite of ratings obtained from both teachers and pupils correlates .80 with the criterion, after attenuation correction. Theoretically, a large battery of tests is necessary for the diagnosis of character; practically, almost as satisfactory results can be

A Comprehensive Study of the Individual 567

obtained simply with ratings and tests of knowledge and opinion about conduct.

No such elaborate investigation is at hand to provide authoritative knowledge concerning personality and its measurement. If by personality is meant an all-round survey of an individual, then a wide variety of tests should be employed. A complete physical and medical examination indicates the adequacy with which the bodily machinery is functioning. Intelligence tests should be used as measures of general level of ability. Questionnaires can be used to determine *interests* over a wide variety of human activities. Questionnaires may also be employed for a psychoneurotic survey to determine adequacy of adjustments, and special tendencies such as introversion-extroversion, ascendance-submission, etc. Tests and questionnaires enable one to tap a wide variety of attitudes and opinions toward moral problems, and personal, social, and economic issues. Performance tests give evidence as to conduct in carefully standardized situations. The nature of emotional reactions may be discovered in part by certain of these questionnaires, by physiological measures, and by the free association experiment. Finally, a more detailed investigation of personal problems and adjustments to them can be undertaken by means of the interview and psychoanalysis. In short, for a complete picture of the personality one would have to use the majority of the techniques described in this book as well as measures of physique and intelligence.

If personality is defined more narrowly, as it usually is by psychiatrists, so that it refers to the adequacy of personal and social reactions and adjustments, then the inquiry can be limited somewhat. Intelligence and aptitude tests will help define the general level of ability and special talents or defects. Performance tests can be used to obtain accurate information as to habits of persistence, inhibition, concentration, honesty, etc. Observation may be used to determine certain behavior characteristics either when among groups as in a class-room or singly while in an interview. Probably the greatest usefulness will be found in ratings, the questionnaire, and the interview for obtaining evidence as to adjustments toward the environment, personal evaluation, attitudes toward reality, sexual relationships, morals, and feelings. Finally, probing for dissociated complexes requires a

568 Diagnosing Personality and Conduct

more painstaking technique which involves free association and psychoanalysis.*

What of the future? As this review of research on the techniques for the diagnosis of conduct is finished, it appears that psychology and sociology are on the brink of a period of intensive research directed toward the development and refinement of these methods, the foundations for which were laid years ago in patient theoretical research in the laboratory. The last decade has seen a beginning made toward using these techniques in the practical affairs of life. In the years immediately to follow one can confidently prophesy that intense and devoted research will greatly extend and refine the practical possibilities of these techniques.

What are the directions that this research will take?

1. There will be a movement toward standardization. Instruments now available will be widely used, so that norms can be derived and the instruments can be accurately used to measure amounts of deviation. One must look to the interview, free association, and psychoanalysis as the most fertile fields for producing the suggestions and ideas that will be appropriated by enterprising and gifted social scientists for elaboration and experimentation, with a view to eventual standardization. Certainly there will be a tendency toward the development of group measures.

2. There will be fundamental research concerning the *reliability* of the various techniques. The conditions affecting the results of each method will receive patient investigation. In particular, there is pressing need for more light on the conditions in which one may expect honesty or dishonesty, the effect of informing the subject of the nature of the test under different situations, and the outcomes of involuntary diagnosis. There will be studies made of the adequacy of the results when obtained from individuals of different characteristics, temperaments, types of adjustment, and the like. Fundamental research will also be undertaken in methods of scoring, recording results, and interpreting results.

3. One may expect a vast amount of research on relationships

* An excellent *Guide to the Descriptive Study of the Personality with Special Reference to Use in Psychiatric Cases* has been prepared by G. S. Amsden and is published by the Bloomingdale Hospital Press, White Plains, New York (1924).

A Comprehensive Study of the Individual 569

between variables similar to the work already accomplished by the Character Education Inquiry. There is reason to believe that this study of relationships will pry into every corner of the field and test out all possible relationships under various conditions. Only by such patient inquiry will it be possible to reach fundamental conclusions as to the significance and validity of the various proposed measures.

4. This study of relationships will lead to exhaustive studies of the items to be included in tests, questionnaires, and interviews. At present there is a feeling that in an interview a rather comprehensive survey is required in order to avoid leaving out any important item. But with the accumulation of relationships it will be possible to draw up schedules of the necessary items on which information must be obtained in order to diagnose problems in various correlated fields.

5. The writer believes that the development of various types of *questionnaires* and *rating methods* promises to yield the results which will be of most value in the immediate future. Because of the significance of language in the control of conduct, particularly the higher levels of conduct, it is probable that methods employing language such as the questionnaire will prove of fundamental value. Measures of the environment and of objective observation techniques also bid fair to reward effort expended upon them. Performance tests are fundamental but exceedingly costly to administer.

The movement toward the investigation and development of techniques for the diagnosis of conduct, now reaching larger proportions every day, promises much for the better control of human affairs.

BIBLIOGRAPHY

GENERAL WORKS RELATING TO THE MEASUREMENT AND DIAGNOSIS OF PERSONALITY AND CONDUCT

- Character Education*, Report of the Committee on Character Education, National Education Association, United States Bureau of Education Bulletin, No. 7 (1926). Chapter 5 is the report of the subcommittee on Character Tests and Measurements, E. D. Starbuck, chairman.
- FROEMMING, A. H., "Bibliography of Character Tests and Measurements," *Journal of Educational Research*, 16: 223-226 (1927).
- FURFEY, P. H., "Tests for the Measurement of Non-intellectual Traits," Catholic University of America, *Educational Research Bulletin*, Vol. III, No. 8 (1928), 35 pp.
- _____, "Tests for Personality Traits," *Catholic Education Review*, 25: 614-621 (1927).
- HENDRICK, I., "The Analysis of Personality," *American Journal of Psychiatry*, 8: 535-563 (1928).
- MANSON, G. E., "A Bibliography of the Analysis and Measurement of Human Personality Up to 1926," Reprint and Circular Series of the National Research Council, No. 72 (1926).
- MAY, M. A., and HARTSHORNE, H., "Objective Methods of Measuring Character," *Pedagogical Seminary and Journal of Genetic Psychology*, 32: 45-67 (1925).
- _____, and _____, "Personality and Character Tests," *Psychological Bulletin*, 23: 395-411 (1926).
- MAY, M. A., HARTSHORNE, H., and WELTY, R. E., "Personality and Character Tests," *Psychological Bulletin*, 24: 418-435 (1927); 25: 422-443 (1928); 26: 418-444 (1929); 27: 485-494 (1930).
- ROBACK, A. A., *A Bibliography of Character and Personality* (Sci-Art Publishers, 1927), 340 pp.
- SYMONDS, P. M., "The Present Status of Character Measurement," *Journal of Educational Psychology*, 15: 484-498 (1924).
- WATSON, G. B., "Character Tests of 1926," *Vocational Guidance Magazine*, 5: 289-309 (1927).
- _____, "The Measurement of the Less Tangible Qualities," *Vocational Guidance Magazine*, 4: 281-289 (1926).
- _____, "A Supplementary Review of Measures of Personality Traits," *Journal of Educational Psychology*, 18: 73-87 (1927).

WATSON, G. B., *Year of Research 1927; Some Investigations published between Jan. 1, 1927, and Jan. 1, 1928, Bearing upon the Program of Religious, Educational, and Social Agencies*, Religious Education Association, Monograph No. 4 (1929).

YOUNG, K., "The Measurement of Personal and Social Traits," *Journal of Abnormal and Social Psychology*, 22: 431-442 (1928).

INDEX

- ABEL, T. M., 187
- Ability and interest, 239, 242-245
- Ability
- estimates of, 478
 - free association as measure of, 387-389
 - level of, 567
 - measurement of, 10, 21, 298
 - to choose principles, test of, 277
 - to judge differences in, 99-104
 - to weigh foreseen consequences, test of, 277, 284, 287, 289
- Abstract
- answers, 291
 - associations, 368
 - words, 374
- Abusing for the defense, 469
- Accepted idea of right, deviation from, score, 74, 75
- Accomplishment
- quotient, 333
 - ratio, 333
- Accuracy
- of diagnosis, 20
 - of interviewing, 453, 454, 475-479
 - of observation, increased by practice, 27
 - of questionnaire, 453, 454
 - of testimony, 475, 498
- Achievement
- college, 193
 - prediction of, 244, 388
 - tests, 11
- Acid, lactic, 412
- Acidity, 437-439, 440
- of saliva, factors in, 439
 - of urine, factors in, 439
- Acquaintance
- effect of, on reliability of ratings, 98, 99
 - factor in rating, 108, 109
- Action
- decisiveness of, 339
 - rising, in interviewing, 460, 462-473
- Activities, extra-curricular, 11
- Activity
- interest, 389, 390
 - mental, factor in psychogalvanic reflex, 429
- Activity (*continued*)
- muscular, factor in acidity of urine, 439
- ADAMS, H. F., 101
- Adaptability, 186
- Addition Test, 311, 312
- Adequacy of sampling, 5
- Adjectives, 374
- descriptive, 50
- Adjustment, 188, 401, 456
- inadequacy of, 567
 - questionnaire, 12, 174-214
- ADLER, A., 486, 492, 493
- Administration
- of rating, 44-50
 - of the questionnaire, 128-130
 - rating as an aid to, 41, 42
- Adrenal glands, 440
- Adrenin, or adrenalin (proprietary name), 400, 401, 405
- factor in blood pressure, 413
- Advanced Bible Knowledge Test, 266
- Advantages
- of graphic rating method, 65
 - of interview, 451-454
 - of rating, 41-44
- Affective tone, 377
- Affectivity score, 161, 191, 192
- Age, 562-564
- educational, 333, 334
 - mental, 321, 334, 351, 563
 - norms, 169, 374
 - variations with, of
 - blood pressure, 411
 - breathing, 418, 419
 - moral knowledge, 288
 - psychoneurotic inventory, 186
 - pulse, 405
 - reaction time, 375
 - testimony, 475
- Aggressiveness, measures of, 331-333
- Agreement
- in the interview, 466
 - of two observers, 34
- Alcohol, 27, 375
- Alertness, essential to observation, 25
- Alkaline tide, 439
- Alkalinity, 437-440
- ALLPORT, F. H., 99, 110, 111, 120, 171, 183, 206, 208, 228, 233, 234, 372, 374, 382

- ALLPORT, G. W., 183, 206, 208
 Alpha, Army, 539-541
 Alternating current, factor in psychogalvanic reflex, 428
 Altruism in rating, 110
 Ambivert, 170, 201
 American Council Intelligence Test, 187, 202, 508
 Amino-acid nitrogen, 438
 Amount, scores of, 161, 306, 311
 Amplitude
 of breathing, measurement of, 417-420
 factors influencing, 418-420
 of heart-beat, measurement of, 407
 pulse, 415
 Analysis
 character, 516
 in item selection, 147, 148
 of personality, 82
 qualitative, of free association, 384
 ANDERSON, V. V., 42
 ANDRUS, R., 31
 Aneroid gage, 409
 Anonymous questionnaire, 453
 Answer, personal or distant, 291
 truthfulness of, 124
 Answering every item, 141, 142
 questionnaire, emotion in, 139
 Appeal of questionnaire, 130
 Apperception Test—see Burdick
 Application, 335
 Applications Test, 275-277, 285-288, 290
 Applied psychologists, 503
 Appointment in the interview, 459
 Aptitude tests, 567
 A R, 334
 Arguments Test, 219, 221
 Arithmetic Problems Test, 311, 312
 ARLETT, A. H., 100
 Army Alpha, 539-541
 Army Rating scale, 57-61, 65, 111, 114
 Arrangement of question blank, 131
 Artery, 408, 409
 brachial, 409
 radial, 403, 404, 409
 Ascendancy-Submission, 567
 Reaction Study, 205-208
 reliability of, 207
 validity of, 207
 ASCHAFENBERG, G., 361, 368, 371
 Assayer, 226
 Association,
 free, method, see Free association
 Associations,
 abstract, 368
 clang, 368, 374
 contiguity, 369
 continuity, 369
 Associations (*continued*)
 controlled, 373
 extrinsic, 368
 histrionic, 368
 inner, 368, 374
 intrinsic, 368, 369
 objective, 370
 of children, 372
 outer, 368, 374, 376
 sound, 368, 369
 subjective, 370
 unpleasant, 375
 Asthenic type, 509, 511-513
 Astrology, 503, 504
 ATHEARN, W. S., 56, 302
 Athletic
 performance tests, 313, 562, 563
 type, 509-511, 513
 Attention
 essential to observation, 25
 factor in free association, 371-373
 Attitude
 measures of, 11, 154, 302
 questionnaires, 12, 215-238; 567
 reliability of, 231-233
 validity of, 233-237
 Self-Rating Scale, Heidebreder Inferiority Attitude, 509
 sex differences in, 235
 social, measures of, 11
 Questionnaire, 157, 158, 166, 215, 216
 toward Movies, Scale of, 229, 230
 Attitudes, 154, 170
 distribution of, 233
 racial, 223
 Test of Social, 167, 224, 225
 Test, Personal, *see* Personal
 training in, 235
 Auscultatory method, 409
 Ausfrage method, 394
 Aussage
 association method, 391, 392
 method, 391
 Average
 boy, criticism of, score, 74, 75
 deviation of reaction-time, 384
 error, in rating, 48-50
 reaction-time, 384
 validity of, 152, 153
 Background
 cultural, 537, 552, 553, 562-564, 566
 home, 536, 550, 552
 Bacterial activity, factor in acidity of saliva, 439
 BAIN, R., 236
 BALDO, C., 525
 Ballot, 215, 237

- Bank, rating used in, 47
 BARKER, M., 35, 39
 BARNES, E. J., 478
 Barometric pressure, factor in breathing, 419
 BARR, F. E., 542, 545
 Barr Scale of Occupational Status, 542-545
 BARRETT, M., 76
 Baths, factor in blood pressure, 412
 Bean Relay Test, 313, 314
 Beauty, 522, 529
 Behavior
 consistency of, 562
 emergency, 400
 observation of, 17
 Problem Record, 68
 problems, 186
 Rating Scale, 69
 BELLINGRATH, G. C., 506
 BEMMELS, V., 308, 310
 BENUSSI, V., 432, 434, 435
 Bericht, 391
 BERNREUTER, R. G., 208
 Personality Inventory, 208
 BERTHOLON, P., 420
 Bible Knowledge Test, Advanced, 266
 Biblical Information, Laycock Test of, 266
 Biblical knowledge tests, 264-267
 Bibliographical research, 126
 Binaural stethoscope, 409
 BINET, A., 15, 267, 268, 329, 351, 526
 Binet-Simon Intelligence Scale, 15, 267, 268, 351
 Binet Vocabulary Test, 329
 BINSWANGER, L., 428
 Biochemical measurements, 437-440
 BLACKFORD, K. M. H., 516, 517, 519, 522, 523, 529, 530
 Blank,
 Interest Analysis, 248, 249
 Interest Report, 166, 248
 question, see Question
 rating, 44, 45
 Student Information, 226
 Vocational Interest, 169, 250, 253
 Blond(e)-brunette types, 504, 522, 523
 Blood,
 alkalinity and creatinine concentration of, 440
 carbon dioxide in, factor in breathing, 419
 pressure, 379, 401, 402, 407-415, 432-434
 diastolic, 410
 factors in, 411-413
 measurement of, 407-414
 systolic, 409, 410, 432-434
 Blood,
 pressure (*continued*)
 variations of breathing with, 418
 of pulse with, 406
 vessels, elasticity of, factor in blood pressure, 411
 volume, changes in, 417
 factors in, 416
 measurement of, 414-417
 variations of blood pressure with, 411
 Bodily characteristics, 514-524
 Body resistance, factor in psychogalvanic reflex, 427
 BOGARDUS, E. S., 468
 Books Read Test, 310
 BOYCE, A. C., 62, 78, 96, 99, 112
 Boy, criticism of average, score, 74, 75
 Boy Scout Organization, 302
 Brachial artery, 409
 BRANHAM, V. C., 289
 Breaking defense mechanisms, 468, 469
 Breathing, 379, 400, 401, 432-434
 factors influencing rate and amplitude of, 418-420
 measurement of amplitude and rate of, 417-420
 variations of blood pressure with, 411, 413
 psychogalvanic reflex with, 428
 BREUER, J., 144
 Bridge, Wheatstone, 422, 423
 BRIDGES, J. W., 186, 187, 192, 193, 242, 243, 329
 BRIDGES, K. M. B., 186, 192, 193
 BRISLEY, M. S., 462
 Broad jump, standing, 313
 BROGAN, A. P., 279, 288-290, 293
 BRONNER, A. F., 289, 323
 BROWN, L. E., 528
 BROWN, W., 33, 95, 162, 201, 253, 307, 310, 313, 319, 320, 325
 BROWN, W. M., 328, 329
 Brown-Spearman formula, see Spearman
 Brown tube, 424
 Brunette, blond(e)-types, 504, 522, 523
 BRUNSWICK, D., 401
 BRYANT, E. K., 349, 351
 BUCKINGHAM, B. R., 333
 BÜHLER, K., 394
 BURDICK, E. M., 552, 563, 564, 566
 Burdick Apperception Test, 552, 563, 564, 566
 reliability of, 552
 BURTON-OPRIZ, R., 411, 418
 BURTT, H. E., 435
 Button, 404

- CADY, V. M., 99, 100, 103, 108, 162,
164, 179, 185, 186, 280, 281, 291.
302, 304, 306-308, 310, 462
- Cady, Woodworth, questionnaire, 164
- Caffeine, 375
- CALDWELL, O. W., 204
- Cancellation Test, 311, 312
- Capacity, judicial, 99, 100, 103
- Capillary, 408
resistance, factor in blood pressure,
411
- Carbon-dioxide, 400, 412, 413
dioxide in air, factor in breathing,
419
dioxide in blood, factor in breath-
ing, 419
- Card, score, see Score
- Cardboard Test, 305, 306
- Carefulness, 339
- CARPENTER, F. L., 350
- Case
conference, 560
history, 494
method, 555
study, 555-560
of development, 560
schedule, 15, 16, 556, 557
worker, 555, 560
- Castration complex, 489
- Catatonic stupor, 431
- Cathode ray, 424
- Cathode-ray oscillograph, 424
- CATTELL, J. McK., 99, 103, 104, 107,
109, 110, 373
- Cause and Effect Test, 271, 277, 285-
288, 290
- Causes,
of crime, 8
of insanity, 10
- Caution, measures of, 328, 329
- CAVAN, R. S., 452
- CAVD Intelligence Test, 324
- CAVI Intelligence Test, 564
- CEI, see Character Education Inquiry
tests, 277, 284, 286
- Censor, 488
- Census, 537, 538
- Certain ratings, 103
- Certainty of opinion, 233, 234
- CHADDOCK, R. E., 131
- CHAMBERS, O. R., 193-195
- CHAPIN, F. S., 546, 547
- CHAPMAN, J. C., 324, 335, 336, 548,
550, 554
- Character, 11, 12, 353, 504, 526, 560,
564, 565
an abstraction, 561
an average, 566
analysis, 516
and consistency, 565
- Character (*continued*)
and dependability, 565
and personality tests, 266
diagnosis of, 560-566
Education Inquiry, 124, 155, 261,
264, 270, 302, 306, 316, 324,
325, 327, 550, 552, 562, 566,
569
estimates of, 566
from photographs, 521, 522
integration of, 566
measures of, 561, 562, 566
portraits, 562, 566
qualities, 11
ratings, 49, 521, 527, 528
reading, 528
traits, 518, 524
Chart, rating, 63, 64
- CHARTERS, W. W., 450, 451, 455
- CHASSELL, E. B., 277, 284, 287,
289
- CHASSELL, J. O., 145, 146, 181, 182,
185, 188
- Chassell Test of Religious Ideas In-
cluding the Ranking of Selected
Answers, 267
- CHASSELL-COOPER, C. F., 267, 268, 277,
284, 287, 289
- CHASSELL-TOORS, L. M., 267, 277, 284,
287, 289
- Cheating, 290, 299
tests of, 303-314
- Cheek list, 31, 72, 73, 563, 566
in observation, 31, 39
reliability of, 73
- Cheeking
in questionnaire, 134, 135
number, test, 311, 312
sheet, in rating, 46
- Children
associations of, 372
feeble-minded, 388
gifted, 243, 389, 545
problem, 9
diagnosis of, 17
- Chinese, 109, 110
- Chinese Ring Test, 324
- Chinning, 313
- Chiromnomy, 524
- Choice of questions in the interview,
471, 472
- Choose Principles, Test of Ability to,
277
- Christianity, 265
- Chronological age, 562-564
- Chronoscope, 365
- Chu, J. P., 591
- Church School Textbooks, Score Card
for Measuring, 56, 57
- Circles Test, 307

- Citizenship, 562-564
 diagnosis of, 8, 11, 12
 ratings as measure of, 49
 Clang associations, 368, 374
 CLARK, E. B., 476-478
 CLARK, W. W., 279-291
 Clark Rating Scale of Offenses, 279, 281
 CLARKE, E. L., 30
 Classes in rating scale, number of, 79, 80
 Classical investigation, 134
 Classification
 of mental disease, 10
 of occupations, 537-539
 counts, 538, 539
 of response in free association, *see* Free association
 Classroom observation, seating plan in, 39
 CLEETON, G. U., 43, 61, 98, 517-519
 Climax in the interview, 460, 473, 474
 Clinical
 experience, 6
 worker, 6, 7
 CLOTHIER, R. C., 61
 CLOTTA, G., 430
 Code, 291-293
 Coefficients of correlation, 20, 21
 reliability, of questionnaire, 164-167
 COGAN, L. C., 521
 Coil, moving, galvanometer, 421, 422
 Coin-Counting Test, 314
 Cold, factor in blood pressure, 413
 Colgate Mental Hygiene Test, 164, 185, 187, 203
 Colgate Personal Inventory, 158, 165, 180, 196, 201, 202
 College achievement, 193
 Color-blindness, 25
 Colorimetric method, Folin, 438, 439
 Combination method, 392
 Combining
 incomplete rankings and ratings, 92, 93
 order-of-merit judgments, 86, 87
 Commercial interest, distribution of, 251
 Commitment, 8, 9, 10
 Committee on Social and Religious Surveys, 302
 Common reactions, 370, 371, 385, 391
 COMMONS, J. R., 547
 Comparison of rating methods, 76-78
 Comparisons, paired, *see* Paired
 Compensation, 491
 Competence, vocational, diagnosis of, 8, 10, 11
 Completion
 Sentence, Test, C. E. I., 311, 312
 Test, Trabue, 329
 Test, Voelker, 303, 304
 Complex, 361, 364, 376, 384, 394, 488, 489, 498
 castration, 489
 constellation, 370
 Electra, 489
 emotional, 432, 433
 Oedipus, 489
 signs, 378-381, 496
 sex differences in, 380
 Composite of ratings, 114
 Comprehension of proverbs, fables and parables, tests of, 267, 268
 Comprehensions Test, 272-276, 285-288, 290
 Comprehensive study of the individual, 555-569
 Concave type, 519, 520
 Conclusion in interviewing, 460, 474
 Concrete words in free association, 374
 Conditions
 of observation, 36-38
 of the interview, 459, 460
 Conduct
 and language, 18
 and moral knowledge, 289, 290
 clinical approach to diagnosis of, 6
 codes, 293
 experimental approach to diagnosis of, 5, 6
 external signs of, 12, 13, 503-535
 frontal attack in diagnosis of, 13, 14
 group, 292, 293
 knowledge, tests of, *see* Knowledge
 measurement of, 4, 5, 7, 298
 principles as a control of, 291, 292
 rating, 52
 record, 563, 564
 results of,
 description of, 12, 13
 measures of, 12
 standards of, 19, 20
 sex differences in, 290
 Contact, 22, 329
 Conference, case, 560
 Confession, 468
 Confidence in rating, 103
 CONKLIN, E. S., 78, 165, 166, 200-203
 Consensus of judgment in determining scoring key, 157, 158
 Consequences
 Foreseen, Test of Ability to Weigh, 277, 284, 287, 289
 Word, Test, 270, 277, 286, 288

- Conservatives, 234
 Consistency,
 and character, 565
 in observation, 34
 of behavior, 562
 of habits, 35
 of measures, 20, 21
 of questionnaires, 149
 of rating, 102, 103
 Constellation, 370, 394
 Contact,
 initial, in the interview, 459
 of electrodes, security of, factor in
 psychogalvanic reflex, 427
 physical, in the interview, 467
 social, observation of, 36
 Contiguity associations, 369
 Continuity associations, 369
 Contradiction, reaction to, 340, 341
 Control,
 Eye, Test of, 331, 332
 of situation, in observation, 36,
 37
 Controlled
 association, 373
 response, 299
 situations, 299
 Convex type, 519, 520
 Cooperation, tests of, 325-327, 563
 Coordination
 of impulses, 342, 343, 346
 Test, 563, 566
 Copying tests, 303
 Correlation, 124
 coefficients of, 20, 21
 used in selecting items for obser-
 vation, 31, 32
 Correlations, small, value of, 559
 Counselor, 14, 15, 476
 vocational, 11
 Counts, G. S., 338, 548
 Counts Classifications of Occupations,
 537-539
 COWDERY, K. M., 147, 148, 159, 166,
 169, 242, 248, 249, 253, 255
 Cowdry Interest Report Blank, 166
 COX, K., 103
 CRANE, H. W., 373, 374
 CRATHORNE, A. R., 239, 240
 Creatinine, 438-440
 concentration of blood, 440
 concentration of urine, factors in,
 439, 440
 CREPIEUX-JAMIN, J., 526, 527
 Crime, 382-385, 433, 436
 causes of, 8
 diagnosis of, 8, 9
 fact of, diagnosis of, 8
 incipient tendencies toward, diag-
 nosis of, 8
 Crime (*continued*)
 prognosis of, 8
 Criminal, 8, 9, 193, 382, 432, 433, 512,
 561
 type, 514-516
 Criterion, 160, 161
 Criticism of the average boy score,
 74, 75
 Cross
 on-a-line method of rating, 62
 -out tests, 189-195, 217, 270,
 281
 Test, 324, 563
 Cultural and Mechanical Environment,
 Questionnaire on, 553
 Cultural background, 537, 552, 553,
 562-564, 566
 Cumulative record, 46
 Current
 electrical, 421, 427
 factor in psychogalvanic re-
 flex, 427, 428
 exsomatic, 423
 CUSHING, H. M., 186, 279, 289, 308,
 310
 Cycloid temperament, 510, 511
 Cyclothymic, 510
 D'arsonval galvanometer, 421, 422
 DASHIELL, J. F., 242, 243, 350
 Data
 for research, ratings provide, 43,
 44
 sheet, personal, 174
 Pupil, 150
 DAVIS, K. B., 453
 DAVIS, W. W., 153, 154
 Deccit, 353, 367
 detection of, 431-437
 measures of, 303-318
 reliability of, 316, 317
 intercorrelations of, 317
 Decisiveness of action, 339
 Decision, speed of, *see* Speed
 Defense, 491
 mechanisms, breaking, 468, 469
 Definition
 in free association, 388
 of qualities in rating, 84-86, 114
 objective, 85
 subjective, 85
 of the problem in questionnaire
 studies, 127, 128
 of what is to be observed, 30
 Degree of truth test, 217
 Delinquency, 13, 193, 349, 350
 diagnosis of, 17
 test of, 390
 DEL MANZO, M. C., 348, 351

- Dementia praecox, 386
 Dénouement in interviewing, 460, 473, 474
 Dependability and character, 565
 Deportment marks, 563, 564, 566
 Deposition, 391
 Description, 19
 of item to be observed, 32
 of reactions, 12, 13
 of the environment, 12, 13
 of the results of conduct, 12, 13
 Descriptive adjectives, 50
 Design of question blank, 130-132
 Desirable traits, 101, 109, 110
 Detection of deceit, *see* Deceit
 of guilt, 361, 382-385
 Detector, lie, 436
 Determination of scoring key, 157, 158, 249, 250
 Development,
 case study of, 560
 observation of, 34, 38
 Deviation,
 from accepted idea of right score, 74, 75
 DeWick, H. N., 420, 425, 430
 Diagnosis, 3, 4
 accuracy of, 20
 of character, 560-566
 of citizenship, 8, 11, 12
 of conduct, clinical approach to, 6
 experimental approach to, 5, 6
 frontal attack on, 13, 14
 of crime, 8, 9
 of delinquency, 17
 of fact of crime, 8
 of incipient tendencies toward crime, 8
 of insanity, 8-10, 17
 of problem children, 17
 of vocational competence, 8, 10, 11
 psychoanalysis as a technique of, 485
 reliability of, 20-22
 verbal vs. non-verbal, 17, 18
 voluntary vs. involuntary, 14, 15, 568
 Diagnostic
 interview, 450, 451
 summary in the interview, 480
 techniques, 6, 7, 15
 Diary in observation, 31, 38, 39
 Diastolic blood pressure, 410
 Diet, factor in acidity of urine, 439
 Differences,
 age, *see* Age
 feeling of, score, 74, 75
 in ability to judge, 99-104
 in persons being rated, 98, 99
 Differences (*continued*)
 in reliability in traits being rated, 104-108
 occupational, in introversion, 202, 203
 racial, 171, 187, 350-353
 sex, *see* Sex
 Difficulty of questions, 470, 471
 Difficulty tests, 311, 312
 Digit-Symbol Substitution Test, 312
 Dime, Planted, Test, 314
 Direct question, 142, 143
 vs. indirect attack, 13, 14
 Directed
 observation, 30, 31, 38
 response, 299
 Direction
 of current, factor in psychogalvanic reflex, 427, 428
 of error in rating, 48-50
 scores of, 161
 Directions for free association, 365
 Tests, Woodworth-Wells, 320
 Disadvantages of the interview, 451-454
 Discovery, 3
 Discrimination,
 essential to observation, 27
 test, ethical, 260, 267
 weight, test, 307, 308
 Disease, mental, classification of, 10
 Disguised Intelligence Test, Snedden's, 454
 questionnaires, 14, 142, 143
 Dishonesty and handwriting, 526
 Dissociation, 372, 387
 Distraction,
 fear, tests, 332
 method, 393
 Distribution
 in rating scale, 80
 of attitudes, 233
 of commercial interest, 251
 of introversion-extroversion, 200
 rectangular, 77, 91
 types of, yielded by questionnaires, 170
 DOLLINGER, W. M., 242, 243
 Dots Test, 312
 Double testing technique, 308-312
 Doubtful reactions, 386
 Down, C. E., 100
 Downey, J. E., 329, 337-339, 342-344, 346-349, 351, 352, 526
 Downey Will-Temperament Test, 329, 337-352
 intercorrelations of, 345, 346, 351
 reliability of, 343, 344, 351
 Dreams, 496
 interpretation of, 500

- Dreams (*continued*)
 symbolism in, 496
 Drew, Tests in Religious Education, 268
 Drives, sexual, 487, 488
 Drugs, 27, 371, 376
 factor in psychogalvanic reflex, 429
 variations of pulse with, 405
 Dudgeon sphygmograph, 404
 DUNLAP, K., 290, 292, 293, 365
 Duplicate questionnaire, 128
 Duplicating technique, 303-306
 Duties Test, 271, 276, 285-288, 290
 Dynamometer, hand, 313
 Dysplastic type, 509, 510
- EASTMAN, F. C., 388
- Eating
 factor in blood pressure, 412
 variations of pulse with, 406
- Ebbinghaus, H., 392
- Economic
 Issues, Watson Survey of Public
 Opinion of, 142, 166, 216-222,
 234, 235
 level of home, 536
 radical, 217-221
 socio-, level, *see* Socio-
 socio-status, Sims scale of, *see* Sims
- EGG, M. D., 363
- Educated, 372
 reaction time of, 375, 376
- Education, Health, Tests, 270, 277
- Education, Inquiry, Character, *see*
 Character
- Education, religious, *see* Religious
- Educational age, 333, 334
- Effect, Cause and, Test, 271, 277, 285-
 288, 290
 halo, 107, 111-113
- Effort in school, measures of, 333-336
- EKDAHL, A. G., 375
- Elasticity of blood-vessels, factor in
 blood pressure, 411
- Electra complex, 489
- Electrical
 current, *see* Current
 resistance, 423, 427
- Electrode, 424, 425, 427
 security of contact of, factor in
 psychogalvanic reflex, 427
 temperature of, factor in psycho-
 galvanic reflex, 428
- Emergency behavior, 400
- Emotion, 376-379, 382, 384, 421, 429-
 431, 436, 488, 489, 496, 498, 517
 in answering questionnaire, 139
 variations with,
 of blood pressure, 413
 of breathing, 419
- Emotion,
 variations with (*continued*)
 of psychogalvanic reflex, 429
 of pulse, 406
- Emotional
 complex, 432, 433
 excitability, 437, 438, 440
 instability, 191
 maladjustment, 194
 maturity, 193, 194
 reactions in the interview, 467, 468
 stability, 562-564
 stats, 393
 susceptibility, 437
 tension, 394
- Emotionality, 187, 192
 score, 191
- Emotions
 Group Scale for Investigating
 (Presscy), 189, 192
 physiological measures of, 12, 400-
 449, 567
 use of, 430-437
- Employee, 42
- Energy of heart, factor in blood pres-
 sure, 411, 413
- ENG, H., 407, 415-417, 420
- Envelopes Test, 326, 562, 563
- Environment, 13
 description of, 12, 13
 home, measure of, 146
 measures of, 12, 536-554, 569
 Questionnaire on Cultural and Me-
 chanical, 553
 questions of, 146
- Epileptics, 387
- Equipment of the living-room, measur-
 ing, 546, 547
- ERLANGER, J., 410
- Erogenous zones, 487
- Error
 average in rating, 48-50
 direction of, in rating, 48-50
 in observation, 24, 498
 of accuracy in testimony, 475,
 498
 of interpretation in psychoanaly-
 sis, 498
 of interpretation in testimony, 475-
 477
 of memory, 498
 of omission in testimony, 475
- Escape, 471
- Essentials to observation, 24-30
- Estimates, 26
 of ability, 478
 of character 521, 522, 566
- Ethical
 Discrimination Test, 260, 267
 Perception Test, 268-279

- Ethical (*continued*)
 -Social Vocabulary Test, Schwezinger, 270, 285, 286, 288
 standards, 290, 291
 Etiquette, 552
 Evaluation of psychoanalytic technique, 498-501
 EVANS, A. L., 519, 520
 Evasions in the interview, 468
 EVERETT, E. M., 187
 Evidence, types of, 12
 Examination
 health, 556
 Illinois Intelligence, 333
 medical, 567
 oral, 478
 physical, 567
 psychiatric, 17
 Excitability, emotional, 437, 438, 440
 Excitement, 475
 freedom from, essential in observation, 30
 Executives, 253
 Exercise
 breathing with, 419
 variations of pulse with, 405, 406
 Experience, 391
 clinical, 6
 variables record, 145, 146, 181, 182, 185
 Experimental
 work in free association, 366
 Experimentalist, 7
 Expiration-inspiration ratio, 435
 Expression
 extreme, 141
 facial, 13
 Exsomatic current, 423
 External signs of conduct, 12, 13, 503-535
 Extra-curricular activities, 11
 Extreme
 opinions, 233, 234
 ratings, 103, 114
 Extrinsic associations, 368
 Extroversion, *see* Introversion
 Extrovert, 170, 171, 201, 234
 Eye-control, test of, 331, 332
 Eyes, 24, 25

 Fables, comprehension of, tests, 267, 268
 Facial profile, 504, 519, 520
 Fact
 of crime, diagnosis of, 8
 score, 306, 311
 Factor, acquaintance, in rating, 108, 109
 generosity, in rating, 96-98

 Factors influencing
 acidity of saliva, 439
 acidity of urine, 439
 blood pressure, 411-413
 blood volume, 416
 creatinine concentration of urine, 439, 440
 psychogalvanic reflex, 426-429
 rate and amplitude of breathing, 418-420
 reliability of rating, 98-108
 response in free association, 371-373
 Facts, questionnaires to obtain, 123, 125-138
 Faculties, 505
 Faculty psychology, 22
 Fair-mindedness, measurement of, 161, 216, 217
 Fallibility of observation, perception, inference, judgment, 15, 16
 FARR, C. B., 512
 FATHERSON, H. F., 509
 Fatigue, 27, 371, 376, 517
 factor in blood pressure, 412
 factor in psychogalvanic reflex, 428
 Fear, 492
 distraction tests, 332
 FECHNER, G. T., 52
 Feeble-minded, 28, 380
 children, 388
 Feeling, 223, 224
 of difference score, 74, 75
 Feelings of inferiority, 493
 FEINGOLD, G. A., 334
 FELEKY, A. M., 435
 FÉRÉ, C., 420, 421
 FÉRÉ's phenomena, 421
 FERENCZI, S., 486
 FERNALD, G., 268, 279, 287, 322, 323
 Fifteen Puzzle Test, 307
 FILER, H. A., 66
 FILTER, R. O., 329, 331
 Finality of judgment, 341, 343, 346
 Finding observation, 30, 31, 38, 39
 Five Rules Test, 272
 Fixation, 487, 489
 Flexibility, 340, 343, 346
 Fluidity of response, 339
 FOLIN, O., 438
 Folin colorimetric method, 438, 439
 Follow-up letters, 129, 130
 Food, factor in creatinine concentration of urine, 439
 Forcefulness, 339
 Foresights Test, 273, 274, 277
 Formal titration method, 438
 Forms, tabulation, 131
 FRANKLIN, E. E., 239-241

- FRANZEN, R., 74, 270, 272, 277, 310, 333, 334
 Free association, 12, 14, 361-399, 430, 431, 485, 495, 496, 498-500, 567, 568
 and insanity, 385-387
 as measure of ability, 387-389
 as measure of interest, 389-391
 classification of responses in, 368-371, 388
 reliability of, 372
 definition in, 388
 experimental work in, 366
 factors conditioning the response in, 371-373
 instructions for, 365
 key for scoring, 390
 list, group method of giving, 364
 List, Jung, 363, 364
 List, Kent-Rosanoff, 362, 363, 377, 381, 385, 386, 388, 390
 qualitative analysis of, 384
 recording time in, 364
 Freedom
 from excitement, essential in observation, 30
 from habits of interpretation, essential in observation, 29
 from load, 339, 340, 343, 346, 347
 from pathological states, essential in observation, 27
 from prejudice, essential in observation, 29, 31
 FREEMAN, F. N., 337
 Freudian, 490, 491, 498
 FREUD, S., 144, 145, 486-488, 492-495, 497-500
 FREYD, M., 65, 66, 86, 147, 159, 183, 197, 242, 246-248, 252, 350, 390
 FRYER, D., 239, 241-244, 540, 541
 Functional mental disorder, 431
 FURFEY, P. H., 95
 Furnishings of the home, inventory of, 546
 FURST, E., 362

 GALL, F. J., 504, 505
 GALLUP, G. H., 94
 GALTON, F., 52, 78, 361, 362, 368
 Galvanometer, 421-424, 427, 430, 431
 deflections, 376, 377
 Games, 313
 GARRETSON, O. K., 250, 251, 253, 255
 Garretson Interest Questionnaire, 250, 251
 reliability of, 252, 253
 validity of, 253-256
 GARRETT, H. E., 93, 187, 193, 508
 GATES, A. I., 263
 Gates-Strang Test of Health Knowledge, 263
 Gage, aneroid, 409
 General trait, 107
 Generalization, 7
 Generalizations Test, 217, 218
 Generosity factor in rating, 96-98
 Gesture, 473
 GINSON, S. M., 329
 Gifted children, 243, 389, 545
 California survey of, 104, 389
 GILBERT, J. A., 319
 GILDEMEISTER, M., 43
 GILES, J. T., 266
 GILLILAND, A. R., 331, 332
 Glands, 510
 adrenal, 440
 pituitary, 440
 sweat, 400, 421, 425-427
 Glandular secretion, 440
 Glasses, 25
 GODEFROY, J. C. L., 423
 GOETHE, J. W., 525
 GOODENOUGH, F., 33, 36
 Good judge, characteristics of, 100, 101
 GORING, C., 515, 516
 GOWIN, E. B., 506
 Graphic rating
 method, advantages of, 65
 scale, 45, 62-72, 76, 77, 159, 180, 196, 197
 reliability of, 95
 rules for making, 65, 66
 Graphology, 525-528
 Gravity, factor in blood pressure, 411
 GRAY, W. S., 459
 GRIFFITHS, C. H., 456
 Grip, 379
 Group
 conduct, 292, 293
 method of giving free association list, 364
 reactions, 372, 374
 Scale for Investigating the Emotions (Pressey), 189, 192
 Growth, observation of, 34, 38
 Guess Who Test, 73, 74, 563, 564, 566
 reliability of, 73, 74
 Guidance, 6, 14, 15
 vocational, 516
 GUILFORD, J. P., 193
 Guilt, 394
 detection of, 361, 382-385
 test of, 382-385
 GULLETTE, R., 187, 192, 377, 434
 GUNDLACH, R. H., 303, 509, 512, 513
 GUTHRIE, E. R., 165, 199, 201, 202, 312
 GUTZMANN, H., 393

- Habit, 292, 560
lists, 31
- Habits,
consistency of, 35
nervous, 31, 35
of interpretation, freedom from
essential in observation, 29
- HAGGERTY, M. E., 68
- HAINES, T. H., 268, 289
- Halo effect, 107, 111-113
- HAMILTON, G. V., 454
- Hand
dynamometer, 313
shape of, 504, 524
- Handwriting, 13; 318-342, 504, 525-
529
and dishonesty, 526
and sex, 526, 529
- HANNA, J. V., 107
- HANSON, W. L., 266
- HARPER, M. H., 167, 232, 235
- HART, H. N., 110, 141, 224-227
- HARTMAN, D. A., 228
- HARTMAN, R., 242, 243
- HARTSHORNE, H., 5, 20, 72, 73, 113,
150, 153-156, 270, 271, 285-293,
298, 299, 302-307, 310-314, 316,
317, 325, 327, 353, 354, 521,
558, 561-566
- HAYES, H., 461
- HAYES, M. S., 94
- Hcad, measurements, 518, 519, 529
- Health
Education Tests, 270, 277
examination, 556
Knowledge, Gates-Strang Test of,
263
- HEALY, W., 268, 287, 289, 474
- Hearing, 412
- Heart-beat
measurement of amplitude of,
407
measurement of rate of, 403-407
- Heart, energy of, factor in blood pres-
sure, 411, 413
- HEIDREDER, E., 147, 165, 170, 196,
197, 200-202, 204, 508, 509
- Heidbreder inferiority attitude self-
rating scale, 509
- Height, 504, 506, 507, 528, 529
- HEINLEIN, C. P., 28
- Helpfulness, 353
tests, 302
- Henriques-Sorenson method, 438
- HEPNER, T. W., 72
- HERAKOWITZ, M. J., 348
- HESKETH, F. E., 511, 512
- HIGHTOWER, P. R., 289
- History, case, 494
- Histrionic associations, 368
- HOFFMAN, G. J., 110, 111
- HORTSMA, R. K., 164, 165, 185, 187,
201, 202
- HOLLEY, C. E., 548
- Hollerith tabulating machine, 131,
132
- HOLLINGWORTH, H. L., 21, 99, 100, 101,
103-107, 109, 110, 112, 174, 186,
187, 478, 499, 521
- Home
background, 536, 550, 552
Conditions, Whittier Scale for
Grading, 548, 549
environment, measure of, 146
furnishings, inventory of, 546
measures of, 146, 546-549
neatness of, 548
necessities of, 548
Sims Score Card for Measuring,
547, 548
size of, 548
socio-economic level of the, 536,
546
Test, 313
- Honesty, 18, 561, 564, 568
measures of, 113, 353
school, 562-564
tests, 150, 302-318, 562-564
- Honor Society, National, 12
- HOUSE, S. D., 145, 146, 164, 180, 181,
183, 185, 186, 188
- House, Woodworth-, Mental Hygiene
Inventory, 164, 180
- HOWELL, H. W., 414
- HUBBARD, L. M., 381
- HUBBARD, R. M., 252, 253
- HUGHES, W. H., 85, 94, 99
- HULL, C. L., 91, 92, 323, 364, 380,
381, 519-521, 524, 525, 527
- Humor, sense of, 457
- Hunger, 486
- HURLOCK, E. B., 109
- Hydrogen-ion concentration, 437
- Hygiene, Mental, *see* Mental
- Hyper-thyroidism, 507
- Hysteria, 180, 431
- Hysteroid, 180
- Idea of right, deviation from accepted,
score, 74, 75
- Ideal, Self-Ordinary, Rating, 74-76
- Identification, 490
- Idiosyncrasy score, 161, 191-193
- Idiot, 388
- I. E. R. tests, 563
- Illinois Intelligence Examination, 333
- Illusion, size-weight, 318, 319
- Imagery, 368
Scale of Mental, 53

- Imbecile, 388
- Importance, 93
 - social, basis for selecting items in observation, 31
- Improbable achievement technique, 306-308
- Improvement
 - in rating, 114
 - prognosis of, 9, 10
- Impulses, coordination of, 342, 343, 346
- Impulsion, motor, 340, 343, 346
- Inaccuracy in perception, 29
- Incomplete
 - rankings, combining, 92, 93
 - ratings, combining, 92, 93
- Incorrigibility, 307
- Incriminating questions, 137, 138
- Independent ratings, 46, 95, 96
- Index
 - Lie, 150, 151, 315
 - Morphologic, 187, 193, 204, 504, 507-509
 - Studiosity, 334-336
 - Truthfulness, 315
- Indiana Survey of Religious Education, 302
- Indicators, complex, 496
- Individual
 - comprehensive study of, 555-569
 - psychology, 492
 - questionnaire to study, 125
 - rating one, at a time, 80, 81
 - reactions, 370, 371, 385, 391
- Industrial classification of occupations, 537
- Industry, 335
- Infantile reactions, 145
- Inference
 - fallibility of, 15
 - Test, 220-222
- Inferiority, 492, 493
 - Attitude, Self-rating Scale Heidbreder, 509
 - feeling of, 493
 - score, 74, 75
- Information
 - social, 235
 - Test, 311, 312, 562, 563
 - Laycock Biblical, 266
- Inhibition, 353
 - motor, 341, 343, 346, 347
 - tests, 302, 327, 328, 562-564, 566
 - Interrelations of, 327
- Initial
 - contact in interview, 459
 - interview, 474
 - polarization, 425
- Inner associations, 368, 374
- Inquiry, Character Education, *see* Character
 - questionnaire used in original, 125
 - statistical, 6, 18
- Insanity, 28, 361, 387, 394
 - and association method, 385-387
 - causes of, 10
 - diagnosis of, 8, 9, 10, 17
 - kind of, 10
 - manic-depressive, 387, 511, 512
- Insight, social, score, 74, 75
- Inspiration-expiration ratio, 435
- Instability, emotional, 191
- Instructions for free association, 365
- Integration, 562
 - of character, 566
 - of personality, 377
- Intellectual interest, 389, 390
- Intelligence, 168, 195, 235, 244, 288, 321, 329, 387, 522, 529, 561-563
- Examination, Illinois, 333
- occupational standards, 540, 541
- Scale
 - Binet-Simon, 15, 267, 268, 351
 - Pearson's for judging, 53, 54
 - tests, 11, 17, 567
 - American Council, 187, 202, 508
 - Army alpha, 539-541
 - C A V D, 423
 - C A V I, 564
 - Disguised, Snedden's, 454
 - National, 287
 - Thorndike, for High School Graduates, 187, 255, 324, 328, 508
- Intelligent, 372, 387, 475
 - reaction time of, 375
- Interchurch World Movement, 302
- Intercorrelations
 - of deceptive behavior, 317
 - of Downey - Will - Temperament Tests, 345, 346, 351
 - of inhibition tests, 327
 - of measures of deceptive behavior, 317
 - of measures of speed of decision, 330
 - of moral knowledge tests, 286
 - of persistence tests, 325
 - of service tests, 326
- Interest, 11, 170
 - activity, 389, 390
 - and ability, 239, 242-245
 - Blank Vocational, 169, 250, 253
 - distribution of commercial, 251
 - free association as measure of, 389-391
 - in detail, 341-343, 346

Interest (*continued*)

- intellectual, 389, 390
- measurement of, 245-251
- medical, scale, 248, 249
- permanency of, 239-242
- Questionnaire, Garretson, *see* Garretson
- questionnaires, 12, 239-259, 567
- Report Blank, 166, 248
- social, 389, 390
- Test, 224, 225

Interpretation, 16, 17, 28-30

- errors of, *see* Errors
- freedom from habits of, essential in observation, 29
- of dreams, 500

Interrogation, 391

Interval, time, in observation, 33

Interview, 6, 7, 12-15, 18, 450-485, 556, 567, 568

- accuracy of, 453, 454, 475-479
- advantages and disadvantages of, 451-454
- agreement in, 466
- appointment in, 459
- choice of questions in, 471, 472
- climax and dénouement, 460, 473, 474
- compared with questionnaire, 451-454

conclusion in, 460, 474

conditions of, 459, 460

dénouement in, 460, 473, 474

diagnostic, 450, 451

diagnostic summary in the, 480

emotional reactions in, 467, 468

evasions in, 468

initial contact in, 459, 474

introduction in, 460-462

involuntary, 14, 457, 458

leads in, 452

motivation in the, 464

non-verbal elements in, 472, 473

note-taking in, 479

observation in, 37

office, 459

physical contact in, 467

preparation for, 458-460

preview in, 458

privacy in, 459

procedures in, 459-474

process, 479, 480

questions in, 469-472

recording in, 479, 480

reliability of, 475-499

research, 450, 451, 458

resistance in, 458, 464-468

rising action in, 460, 462-473

schedule, 454, 461

single, 474

Interview (*continued*)

standardization of, 6, 454, 455

steps in, 459-474

suggestion in, 452

tension in, 464-468

treatment, 451

voluntary, 457, 458

Interviewer, 15, 27, 455-457

training of, 457

Interviews, types of, 450, 451

Intrinsic associations, 368, 369

Introduction in interviewing, 460-462

Introspection, 394

of children, 143, 144

Introversion, 21, 22

characteristics of, 197-199

-extroversion, 567

distribution of, 200

occupational differences in, 202, 203

performance tests of, 200

questionnaire to measure, 165, 166, 195-205

reliability of, 201

validity of, 202-205

sex differences in, 204

Introvert, 170, 171, 201, 234

Inventory

Bernreuter Personality, 208

Colgate Personal (Laird), 158,

165, 179, 180, 196, 201, 202

of home furnishings, 546

Psychoneurotic, *see* Psychoneurotic

Thurstone Neurotic, 208

Woodworth Psychoneurotic — *see*

Woodworth

Woodworth-House Mental Hygi-

ene, 164, 180

Involuntary

diagnosis, 14, 15, 568

interview, 14, 457, 458

I. Q., 103, 187, 235, 243, 287, 351,

562

Item

answering every, 141, 142

in a rating scale, should be ob-

servable, 84

multiple-choice, 140

occurrence of, in observation, 33

selection, based on a social impor-

tance, 31

based on social selection, 147,

148

by analysis, 147, 148

empirical method of,

for observation, 31, 32

for questionnaire, 147-149

to be observed, description of, 32

true-false, 140

validation by rating, 147

- Item (*continued*)
 yes-no, 140
 number of, in a rating scale, 83
 weighting, *see* Weighting
- JACOBSEN, C., 187, 192, 377
 Japanese Cross Test, 324, 563
 JEFFRESS, L. A., 425, 426
 JOHNSON, B., 179
 JOHNSON, I., 375
 JONES, E. 368
 JONES, E. S., 167, 233, 235
 Judge, characteristics of good, 100, 101
 differences in ability to, 99-104
 Judging intelligence, scale for, 53, 54
 Judgment, 79, 27
 consensus of, in determining scoring key, 157, 158
 fallibility of, 15, 16
 finality of, 341, 343, 346
 in rating, 51
 made analytical by ratings, 43
 made representative by rating, 43
 moral, tests, *see* Moral
 tests of, 12, 260-297
 reliability of, 284, 285
 validity of, 285-294
 variability of, 102
 Judgments, order-of-merit, combining, 86, 87
 Judicial capacity, 99, 100, 103
 Jump, standing broad, 313
 JUNG, C. G., 147, 196, 361-366, 368, 370-376, 379-381, 387, 391, 430, 486
 Jung Free Association List, 363, 364
- KAKISE, H., 394
 KALMUS, H. T., 428
 KELLEY, T. L., 88, 91, 159, 162, 244, 248, 282, 302, 335, 336, 388, 389, 551
 Kelley-Wood Table of the Normal Probability Integral, 88, 91, 551
 KELLOGG, W. N., 187, 508
 KENT, G. H., 361-363, 370, 371, 374, 377, 381, 385-388, 390
 Kent-Rosanoff Free Association List, 362, 363, 377, 381, 385, 386, 388, 390
 Key, *see* Scoring
 KINDER, V. S., 110
 Kinetic Will Test, 322, 323
 KING, L. H., 244
 KINGSBURY, F. A., 47
 Kits Test, 326, 562, 563, 566
 KITSON, H. D., 506
 KLEIN, J., 392, 393
- Knee-Jerk, 379
 KNIGHT, F. B., 43, 61, 74, 77, 97, 98, 108, 112, 113, 310, 517-519
 Knowledge
 and judgment, tests of, *see* Judgment
 and opinion, *see* Opinion
 biblical, tests of, 264-267
 Health, Gates-Strang, Test of, 263
 moral, and age, 288
 and conduct, 289, 290
 moral, tests, *see* Moral
 of conduct, tests of, 12, 260-297, 563-567
 intercorrelations of, 286
 reliability of, 284, 285
 validity of, 285-294
 Test, Advanced Bible, 266
 Word, Thorndike Tests of, 261
- KOHS, C. H., 380, 391, 394
 KOHS, S. C., 260, 267
 KOLL, A. G., 410
 KORNHAUSER, A. W., 94, 95, 529, 548
 KRAMER, F., 391
 KRÄPELIN, E., 361, 368
 KRETSCHMER, E., 504, 507, 509-513, 514
 Kretschmer's types, 504, 509-514
 Kymograph, 392, 402-404, 410, 416-418
- Labor, factor in blood pressure, 413
 Lactic acid, 412
 LAIRD, D. A., 158, 159, 165, 171, 179, 180, 183, 196, 197, 201-204, 208
 LANDIS, C., 51, 98, 100, 187, 192, 377, 420, 425, 428, 430, 432, 434, 435
 LANGFELD, H. S., 375, 382
 LANGLEIE, T. A., 244
 Language and conduct, 18
 LARSON, J. A., 407, 432, 434
 LASHLEY, K. S., 505
 LASLETT, H. R., 390
 LAYCOCK, S. R., 266
 Laycock Test of Biblical Information, 266
 LEACH, H. M., 383
 Leadership, 12, 204, 506, 507, 509, 529
 Leading questions, 135-137, 471
 Leads in interview, 452
 Lehmann plethysmograph, 415, 416
 LEIBNITZ, G. W., 525
 Length of observation period, 37, 38
 LENTZ, T. F., Jr., 147, 193
 Leptosome type, 513
 Letter, follow-up, 129, 130
 preliminary, 130
 Level, *see* Socio-economic, Occupation
 Lever arm, 403
 Libido, 486
 Lie detector, 436

- Lie Index, 150, 151, 315
 LINDBERGH, C. A., 50
 LINDSAY, E. E., 99, 129
 LIPMANN, O., 392
 List, check, *see* Check
 free association, *see* Free association
 habit, 31
 Thorndike Word, 261, 364
 Living-room, measuring the equipment
 of, 546, 547
 Load, freedom from, 339, 340, 343, 346,
 347
 LOMBROSO, C., 432, 504, 514-516
 LONG, J. A., 160
 LOOMIS, A. M., 32
 Loss of reliability, 79
 LOWE, G. M., 268, 289
 LOWELL, F., 372
 LUDGATE, K., 522, 523
 LUDWIG, K. F. W., 402
 LUGOFF, L. S., 364, 380, 381
 Lying, 303, 432, 433
 test of, 382-385
 tests of, 315

 MACALISTER, A., 504, 505
 MACLAURIN, D. O., 524
 Macroplanchic, 507
 Magic
 number square, 324
 Square Test, 314
 word square, 324
 Maladjustment, emotional, 194
 MALLER, J. B., 298, 323, 327
 Man, Mystery, Test, 314
 Maniac, 393
 Manic-depressive insanity, 387, 511, 512
 Manipulation Test, 327
 Manners, Good, Orr Test, 264, 270,
 286, 552
 Manometer, 408, 409
 MANSON, G. E., 104, 110, 328
 MANTEGAZZA, P., 52
 Man-to-man rating, 57-61
 reliability of, 95
 Manual, raters, 47, 48
 MÄRBE, K., 394
 MAREY, E. J., 403, 404, 417, 418
 Marks
 of deportment, 563, 564, 566
 school, 79, 478, 562-564, 566
 MARSTON, L. R., 165, 166, 196, 199-
 201, 204
 MARSTON, W. M., 367, 407, 432-434
 MARTIN, H. E., 372
 MATEER, F., 386
 MATHEWS, E., 164, 178, 185-187, 193
 Mathews, Woodworth-, questionnaire,
 164, 193

 Maturity, emotional, 193, 194
 MAULSBY, W. S., 461
 MAY, M. A., 5, 20, 72, 73, 113, 150,
 153-156, 270, 271, 285-293, 298,
 299, 302-307, 310-314, 316, 317,
 325, 327, 337, 349, 353, 354,
 558, 561-566
 Maze Test, 306, 307
 McCABE, F. E., 521
 MCCALL, W. A., 160, 235, 453
 McCall-Long method of weighting
 items, 160
 MCCURDY, J. T., 174
 McDOWELL, R. J. S., 426
 MCFADDEN, J. F., 350
 MCGROCH, J. A., 164, 165, 191
 Measure-
 adjustment, questionnaire to, 12,
 174-214
 attitude, questionnaire to, 12, 215-
 238, 567
 interest, questionnaire to, 12, 239-
 259, 567
 introversion-extroversion, question-
 naire to, *see* Introversion
 of ability, free association as, 387-
 389
 of character, rating as, 49
 of citizenship, ratings as, 49
 of equipment of living-room, 546,
 547
 of home environment, 146
 of interest, free association as, 389-
 391
 of mental unbalance, 9
 of prejudice, 161
 Measurement, 19, 450
 observation as, 32-36
 of amplitude of breathing, 417-420
 of amplitude of heart-beat, 407
 of blood pressure, 407-414
 of blood volume, 414-417
 of conduct, 4, 5, 7, 298
 of fair-mindedness, 161, 216, 217
 of interests, 245-251
 of personality, 560, 561, 567, 568
 of psychogalvanic reflex, 420-429
 of rate of breathing, 417-420
 of rate of heart-beat, 403-407
 scale of, traits as, 50-52
 Measurements,
 biochemical, 437-440
 head, 518, 519, 529
 physiognomic, 518, 520
 Measures
 consistency of, 20, 21
 of ability, 10, 21, 298
 of aggressiveness, 331-333
 of caution, 328, 329
 of character, 561, 562, 566

Measures (*continued*)

- of deception, *see* Deceit
 - of effort in school, 333-336
 - of emotions, *see* Emotions
 - of environment, 12, 536-554, 569
 - of home, 546-549
 - of honesty, 113, 353
 - of intelligence, *see* Intelligence
 - of occupational level, 537-545
 - of physique, 567
 - of reactions, 12
 - of responses, 536
 - of results of conduct, 12
 - of social attitude, *see* Social
 - of socio-economic level, 548, 550-552, 563
 - of speed of decision, *see* Speed
 - of stiduousness, 333-336
- Measuring
- Church School Textbooks, Score Card for, 56, 57
 - the Home, Score Card for, 547, 548
- Mechanical Ability Study, Minnesota, 545, 553
- Mechanical and Cultural Environment, Questionnaire on, 553
- Mechanisms, 489-491, 497
- breaking defense, 468, 469
 - of mental release, 468
- Medical
- examination, 567
 - interest scale, 248, 249
- MEIER, N. C., 348, 351
- Melancholic temperament, 393
- MELTZER, H., 452, 470
- Memories, 146
- Memory, error in, 498
- MENDELSSOHN, M., 426
- Menstruation, factor in blood pressure, 413
- Mental
- Ability, Terman Group Test of, 351
 - activity, factor in psychogalvanic reflex, 429
 - age, 321, 334, 351, 563
 - balance, prognosis of, 9
 - disease, classification of, 10
 - disorder, 431
 - Hygiene Test, Colgate, 164, 185, 187, 203
 - Hygiene Inventory, Woodworth-Housc, 164, 180
 - Imagery, Scale of, 53
 - Morgan Group, Test, 351
 - release, mechanisms of, 468
 - test, *see* Mental
 - unbalance, degree of, measure of, 9

Mercury manometer, 409

MESARA, A., 394

Metabolism, 402, 438-440

Method

- auscultatory, 409
- ausfrage, 394
- ausage, 391
 - association, 391, 392
- case, 555
- combination, 392
- distraction, 393
- empirical, of selecting items, *see* Empirical
- Folin colorimetric, 438, 439
- formal titration, 438
- free association, *see* Free association
- graphic rating, *see* Graphic
- Henriques-Sorenson, 438
- of making ratings, 52-76
- of paired comparisons, 54, 55, 76
- of rating, cross-on-a-line, 62
- of weighting items, McCall-Long, 160
- order-of-merit, 55, 76
- perception, 392, 393
- reproduction, 391
- scientific, 23
- scentence, 393, 394

Methods

- of scoring questionnaires, 156-161
- rating, *see* Rating
- statistical, 6, 7, 20
- testing, 6, 139-142

M. I. *see* Morphologic Index

Microsplanchic, 507

MILLER, G. F., 315

Mimeographing, 132

MINER, J. B., 85, 94, 106

Minnesota,

College Ability Tests, 508

Mechanical Ability Study, 545, 553

Mirror galvanometer, 421, 422

MOHR, G. J., 509, 512, 513

Money Voting Test, 326, 563

MONROE, R., 459

MONROE, W. S., 333

MONTGOMERY, R. B., 527

MOORE, B. V., 242, 245, 246

MOORE, H. T., 216, 331

Moral

- Judgment Test, 221, 222
- judgment tests, 268-279
- knowledge and age, 288
- and conduct, 289, 290
- knowledge tests, 285-294, 302, 563, 566, 567
- intercorrelations, 286
- reliability of, 284, 285
- validity of, 285-294

- MORAVESIK, M. H., 393, 428
 MORGAN, J. J. B., 323, 351
 Morgan Group Mental Test, 351
 Morphologic Index, 187, 193, 204, 504,
 507-509
 Morphology, 507
 Motivation in the interview, 464
 Motive, 294
 Motor
 impulsion, 340, 343, 346
 inhibition, 341, 343, 346, 347
 Movement
 muscular, factor in psychogalvanic
 reflex, 427
 speed of, 339, 343-346
 Movies, Scale of Attitude toward, 229,
 230
 Moving coil galvanometer, 421, 422
 MOWREN, E. R., 463
 Multigraphing, 132
 Multiple-choice item, 140
 MÜNSTERBERG, H., 25, 136, 361, 364,
 382
 Muscio, B., 137
 Muscle-reading, 338
 Muscular
 activity, factor in acidity of urine,
 439
 movement, factor in psychogal-
 vanic reflex, 427
 theory of psychogalvanic phenom-
 enon, 425, 426
 Mystery Man Test, 314
- NACCARATI, S., 193, 507, 508
 NAPOLEON, 506
 National Honor Society, 12
 National Intelligence Test, 287
 National Research Council, 174
 Natural
 response, 299
 situations, 299, 300
 Neatness, 528, 529
 of the home, 548
 Negative affective tone, 377
 Negative question, 137
 Negro, 350
 Nervous,
 habits, 31, 35
 sympathetic system, 400
 Nested Squares Test, 306, 563
 Neurasthenia, 180
 Neuraathenoid, 180
 Neurosis, 488
 Neurotic Inventory, Thurstone, 408
 Newark Survey, 151, 152
 NEWCOMB, A., 529
 NEWCOMB, T. M., 32
 New-type test, 222
- Nitrogen, amino-acid, 438
 Non-compliance, 341, 343, 346
 Non-verbal elements in the interview,
 472, 473
 vs. verbal diagnosis, 17, 18
 Normo-splanchnic, 507
 Norms, 169, 374
 Norworthy, N., 98, 104
 Note-taking, 47, 479, 494
 preliminary to rating, 46, 47
 Number
 Checking Test, 311, 312
 of items in a rating scale, 83
 of observations, 33
 of questions, 132
 of scale-divisions in rating, 78-80
 of social situations, observation of,
 36
 square, 324
 NUNBERG, H., 379, 380
- Objective
 associations, 370
 definition in rating, 85
 record, rating should be, 45
 test, 222, 556
 reliability of, 168
 trait, 106, 107
 Objectivity in rating, 76
 Observable items in a rating scale, 84
 Observation, 7, 12, 13, 17, 23-40, 298,
 450, 567, 569
 accuracy of, increased by prac-
 tice, 27
 as measurement, 32-36
 conditions of, 36-38
 consistency in, 134
 control of situation in, 36, 37
 defining the objects of, 30
 description of item in, 32
 diary in, 31, 38, 39
 directed, 30, 31, 38
 discrimination essential to, 27
 errors of, 24, 498
 essentials to, 24-30
 fallibility of, 15
 finding, 30, 31, 38, 39
 freedom from excitement, essen-
 tial to, 30
 increased by practice, 27
 increasing reliability of rating, 99,
 108
 in the interview, 37
 occurrence of an item in, 33
 of behavior, 17
 of growth or development, 34, 38
 of number of social situations, 36
 of social contact, 36
 of time spent in social situation, 36

- Observation (*continued*)
 participation in the situation in, 37
 period, length of, 37, 38
 preliminary to rating, 46, 47, 114
 record of, 28
 recording, 38, 39
 reliability of, 31, 33-36
 seating plan in classroom, 39
 selection in, 30
 selection of items in, 31, 32
 self-, 18
 systematic, 49
 time interval in, 33
 unit of measurement in, 33
 Observations,
 number of, 33
 spacing of, 33
 Observers, agreement of, 34
 Occupational
 differences, in introversion, 202,
 203
 intelligence standards, 540, 541
 level, measures of, 536-545
 Preference, Questionnaire of, 245,
 246, 248
 Status, Barr Scale of, 542-545
 Occupations
 classifications of, 537-539
 Counts, 538, 539
 industrial, 537
 Taussig Scale of, 538
 Oedipus complex, 489
 Offense rating, 279-284
 Clark Scale of, 279, 281
 Office interview, 459
 OLANDER, E. 110
 OLSON, W. C., 24, 31, 33-36, 38, 68
 Omission, errors of, in testimony, 475
 Opinion
 certainty of, 233, 234
 extreme, 233, 234
 Public, Watson Survey of, on Some
 Religious and Economic Issues,
 142, 166, 216-222, 234, 235
 tests of, 167, 562-567
 Opposites Test, 270, 285, 286, 288
 Opposition, resistance to, 341
 Optimism in rating, 110
 Oral examination, 478
 Order-of-merit
 judgments, combining, 86, 87
 method of rating, 55, 76
 Ordinary-Ideal, Self-Rating, 74-76
 Organic
 inferiority, 492
 mental disorder, 431
 Organization, verbal, 566
 Organs, sense, 24
 Original inquiry, questionnaire used in,
 125
 O'ROURKE, L. J., 66, 454
 ORR, C. I., 263, 264, 552
 Orr Good Manners Test, 264, 270, 286,
 552
 Oscillograph, cathode-ray, 424
 OTIS, A. S., 267
 OTIS, M., 320, 321
 Outer associations, 368, 374, 376
 Overrating, 74, 109, 110
 Overstatement Test, 308-310
 Pain
 factor in blood pressure, 413
 factor in breathing, 419
 Paired comparisons,
 method of, 54, 55, 76
 statistical treatment of, 86-89, 93
 PALMER, V. M., 463
 Palmistry, 503, 504
 Paper
 and pencil test, 12, 402
 Paraffin Test, 303, 304
 sensitive, 403
 smoked, 402, 403
 Parables tests, 267, 268
 Paraffin Paper Test, 303, 304
 Paranoia, 387
 Parental conditions, 548
 supervision, 548
 Paresis, 387
 PARTEN, M., 33
 Participation in the situation, in ob-
 servation, 37
 PATERSON, D. G., 61, 85, 94, 97, 509,
 522, 523
 Pathological states, freedom from, es-
 sential to observation, 27
 Patient, 493, 494
 psychopathic, 350
 PEARSON, K., 53, 78
 Peg, Puzzle, Test, 307
 Pelotte, 404
 Percentages, 138, 142
 Percentile scores, 169
 Percentiles, turning ranks into, 91, 93
 Perception,
 essential to observation, 28, 29
 fallibility of, 15
 inaccuracy in, 29
 method, 392, 393
 Test, Ethical, 268-279
 Performance tests, 12, 113, 298-366,
 565, 567, 569
 of introversion, 200
 Permanency of interest, 239-342
 PERRY, C. A., 546
 Perseveration, 379, 381
 volitional, 342, 343, 346
 PERSING, K. M., 150, 154

- Persistence tests, 302, 321-325, 339,
353, 562-564, 566
intercorrelations of, 325
reliability of, 325
- Personal
answer, 291
Attitudes Test, 74
reliability of, 76
Data Sheet, 174
Inventory, 179
Colgate, 158, 165, 180, 196,
201, 202
- Personality, 437-440
analysis of, 82
and character tests, 266
integration of, 377
Inventory, Bernreuter, 208
measurement of, 560, 561, 567,
568
Rating Scale, 72
ratings, 519, 520
Report, 70, 71
Schedule, 183-185, 187
- Personnel spirit, 42, 43
- Perspiration, 401
- PETERS, C. C., 56
- PETERSON, F., 376, 377, 428, 430
- Phenomena
Féré's, 421
Tarchanoff's, 421
- Phonograph, 393
- Photographs, 504, 521, 522
estimating character from, 521,
522
- Phrenology, 503-505
- Physical
contact in the interview, 467
examination, 567
- Physiognomic measurements, 518,
520
- Physiognomy, 504
- Physiological measures of the emotions,
see Emotions
- Physique, 504-514, 561
measures of, 567
- Picture inhibition test, 563, 566
- Pituitary gland, 440
factor in blood pressure, 413
- PLACE, M. J., 100, 160
- Planted Dime Test, 314
- Plethysmograph, 415, 416
- Pneumograph, 417, 418
- POFFENBERGER, A. J., 130, 136, 350
- Polarization, 424-427
initial, 425
of electrodes, factor in psychogal-
vanic reflex, 427
- PORTER, E., 266
- Portraits, character, 562, 566
- Positive affective tone, 377
- Posture, 13
factor in breathing, 419
variations of pulse with, 406
- Potassium salts, 405
- Potentiometer, 423
- Practice, 371, 374-376, 378, 475
accuracy of observation, increased
by, 27
- Prediction of school achievement, 244,
388
- Preference, occupational, question-
naire of, 245, 246, 248
- Pregnancy, factor in blood pressure, 413
- Prejudice, 218-220
freedom from essential in observa-
tion, 29, 31
measure of, 161
- Preliminary letter, 130
- Preparation
for the interview, 458-460
of question blank, 130-132
of questions, 132-138
- Presser, S. L., 161, 164, 165, 168, 189,
191, 193, 195, 281, 478, 508
- Pressey X-O Tests 161, 164, 165, 168,
189-195, 508
reliability of, 191, 192
- Pressure
barometric, factor in breathing,
419
blood, *see* Blood
diastolic, 410
of electrodes, factor in psychogal-
vanic reflex, 428
systolic, 409, 410, 432-434
- Preview, in interviewing, 458
- PAXYER, W., 525
- PRIDEAU, E., 320
- Principles
as a control of conduct, 291, 292
Test, 274, 275, 277, 285-288, 290
weigh and choose, 277
- Privacy in the interview, 459
- Probability, 52
integral, normal, Kelley-Wood table
of, 88, 91, 551
- Probable error, weighting the, 148, 149
- Problem
defining, in questionnaire studies,
127, 128
Record, Behavior, 68
- Problem children, 9
diagnosis of, 17
- Problems,
Arithmetic, Test, 311, 312
behavior, 186
- Procedures in interviewing, 459-474
- Process interview, 479, 480
- Profile, 337, 342, 349
facial, 504, 519, 520

- Prognosis
 of erimc, 8
 of improvement, 9, 10
 of mental balance, 9
- Projection, 490
- Protestants, 265
- Proverbs Test, 267
- Provocations Test, 272, 273, 276, 285-288, 290
- Psychasthenia, 180
- Psychasthenoid, 180
- Psychiatric examination, 17
- Psychiatrist, 10, 17, 451, 456, 560, 567
- Psychoanalysis, 6, 12, 14, 144, 145, 485-502, 567, 568
 a technique of diagnosis, 485
 a therapeutic measure, 485
 errors of interpretation in, 498
- Psychoanalyst, 138
- Psychoanalytic
 technique, 493-497
 evaluation of, 498-501
 theory, 485-493
- Psychogalvanic phenomena
 muscular theory of, 425, 426
 sweat theory of, 425, 426
 vasomotor theory of, 425, 426
- Psychogalvanic reflex, 379
 factors influencing, 426-429
 measurement of, 420-429
- Psychological questionnaire, 15, 16, 139-171
- Psychologist, 17
 applied, 503
- Psychology
 faculty, 22
 individual, 492
- Psychoncurotic Inventory, 146, 154, 168, 169, 171, 174-214, 567
 reliability of, 185, 186
 validity of, 186-188
 Woodworth, *see* Woodworth
- Psychopathic
 patients, 350
 states, 394
- Psychopathy, 386
- Psychosis, 512
- Ps, A. S. T., 109
- Pulse, 379, 403-408, 415, 434
 amplitude, 415
 factors in breathing, 418
 rate, 401-407, 415
 variations of, 405-407
- Pupil Data Sheet, 150
- Puzzle
 Fifteen, Test, 307
 Manipulation Test, 327
 Peg Test, 307
- Pyknic type, 509-512
- Qualitative analysis of free association, 384
- Qualities,
 character, 11
 definition of, in rating, 84-86, 114
 to be rated,
 selecting, 81, 82
 weighting, 82, 83
- Quantitative scores, yielded by ratings, 45
- QUEEN, S. A., 473
- Question
 blank
 arrangement of, 131
 design of, 130-132
 preparation of, 130-132
 size of, 130, 131
 difficulty of, 470, 471
 direct, 142, 143
 incriminating, 137, 138
 leading, 135-137, 471
 negative, 137
 subtle, 142, 143
 suggestive, 137, 471
 yes-no, 140
- Questionnaire, 6, 13, 14, 16, 18, 19, 22, 122-173, 556, 569
 accuracy of, 453, 454
 adjustment, 12, 174-214
 administration of, 128-130
 anonymous, 453
 appeal of, 130
 Ascendancy-Submission, *see* Ascendancy
 attitude, *see* Attitude
 checking in, 134, 135
 compared with interview, 451-454
 consistency of, 149
 disguised, 14, 142, 143
 distribution, types of, yielded by, 170
 duplicate, 128
 Garretson Interest, *see* Garretson
 interest, *see* Interest
 introversion-extroversion, *see* Introversion
 Occupational Preference, 245, 246, 248
 on Cultural and Mechanical Environment, 553
 psychological, 15, 16, 139-171
 Psychoncurtic, *see* Psychoncurtic, Woodworth
 rating in a, 141
 reliability of, 161-169
 returns on, 129, 130
 scoring, methods of, 156-161
 selection of items for, 147-149
 sex differences in, 171

- Questionnaire (*continued*)
 social attitudes, 157, 158, 166, 215, 216
 studies, defining problem in, 127, 128
 studiousness, *see* Studiousness
 tabulation in, 135, 136
 to obtain facts, 123, 125-138
 to study individuals, 125
 trying out, 128
 used in original inquiry, 125
 validity of, 147-156
 Woodworth Psychoneurotic, *see* Woodworth
 Woodworth-Cady, 164
 Woodworth-Mathews, 164, 193
- Questions, 436
 in the interview, 469-472
 choice of, 471, 472
 number of, 132
 of environment, 146
 of reactions, 146
 preparation of, 132-138
 should be brief, 132, 133
 should be simple, 133
 should be specific, 133, 134
 should be unambiguous, 133
 should cover information desired, 133
- Quotient, accomplishment, 333
- Racial
 attitudes, 223
 differences, 171, 187, 350-353
- RADEMACHER, E. S., 465, 474
- Radial artery, 403, 404, 409
- Radical, 234
 economic, 217-221
- RANK, O., 486
- Ranking, 55, 76-78, 97, 98, 107
 combining incomplete, 92, 93
 offense, 279, 280
 statistical treatment of, 90-93
- Ranks, turning
 into percentiles, 91, 93
 into standard deviation units, 91, 93
- RANNELLS, M. E., 464, 466
- Rapport, 461, 462
- Rate
 of breathing, measurement of, *see* Breathing
 of heart-beat, measurement of, 403-407
 pulse, 401-407, 415
- Rater's manual, 47, 48
- Raters, training, 47-50
- Rating, 41-121, 347, 348, 350, 450, 518, 562, 565-567
 acquaintance factor in, 108, 109
- Rating (*continued*)
 administration of, 44-50
 advantages of, 41-44
 altruism, in, 110
 as an aid to administration, 41, 42
 average error in, 48-50
 blanks, 44, 45
 chart, 63, 64
 checking sheet in, 46
 conduct, 52
 confidence in, 103
 consistency of, 102, 103
 cross-on-a-line method of, 62
 definition of qualities in, 84-86, 114
 objective, 85
 differences in traits in, reliability in, 104-108
 direction of error in, 48-50
 factors influencing reliability of, 98-108
 generosity factor in, 96-98
 graphic, *see* Graphic
 improvement in, 114
 in a questionnaire, 141
 independent, 46
 judgment in, 51
 man-to-man, *see* Man-to-man
 methods, 6, 12, 13, 41-121, 569
 comparison of, 76-78
 note-taking, preliminary to, 46, 47
 number of scale divisions in, 78-80
 objectivity in, 76
 offense, 279-284
 one individual at a time, 80, 81
 one trait at a time, 80, 81
 optimism in, 110
 order-of-merit method, 55, 76
 paired comparison, 54, 55, 76
 personality, 519, 520
 preceded by observation, 46, 47, 114
 purposes of, 41
 regression in, 104
 reliability of, 93-96
 effect of acquaintance on, 98, 99
 factors influencing, 98-108
 increased by observation, 99, 108
- scale,
 Army, 57-61, 65, 111, 114
 Behavior, 69
 distribution in, 80
 graphic, *see* Graphic
 items observable in, 84
 number of classes in, 78-80
 number of items in, 83
 of offenses, Clark, 279, 281
 Personality, 72

Rating (*continued*)

- schedule, 42
- self-, 109-111, 141, 234
- Self-Ordinary-Ideal, 74-76
- sex difference in, 101, 102
- sheets, 45, 46
- should be objective record, 45
- traits, 52
- used in a bank, 47
- validating test items by, 147
- validity of, 108-113
- variability in, 93

Ratings

- as measure of character, 49
- as measure of citizenship, 49
- certain, 103
- character, 49, 521, 527, 528
- combining incomplete, 92, 93
- composite of, 114
- extreme, 103, 114
- given periodically, 44
- given systematically, 44
- independent, 46, 95, 96
- make judgments analytical, 43
- make judgments representative, 43
- methods of making, 52-76
- provide data for research, 43, 44
- recording, 54
- stimulate the person being rated, 42
- systematic, 114
- yield quantitative scores, 45

Ratio

- accomplishment, 333
- inspiration-expiration, 435

Rationalization, 142, 219, 221, 222, 489

RAUBENHEIMER, A. S., 279, 281, 284, 287, 289, 302, 308, 310

Ray, cathode, 424

Reaction

- Study, Ascendancy-Submission, *see* Ascendancy
- time, 373-378, 381, 384, 394, 430, 431
 - age differences in, 375
 - average, 384
 - average deviation of, 384
 - of educated, 375, 376
 - of intelligent, 375
 - sex differences in, 375, 376
- to contradiction, 340, 341

Reactionary, 234

Reactions

- common, 370, 371, 385, 391
- description of, 12, 13
- doubtful, 386
- emotional, in the interview, 467, 468
- group, 372, 374
- individual, 370, 371, 385, 391

Reactions (*continued*)

- infantile, 145
- measures of, 12
- questions of, 146

Reading

- character, 528
- Test, Thorndike-McCall, 235

REAM, M. J., 91, 242, 246

Reasoning tests, 287

Recognition tests, 274, 277, 285-288, 290

Record

- Behavior Problem, 68
- conduct, 563, 564
- cumulative, 46
- diary, 31
- objective, rating should yield, 45
- of observation, 28

Recording

- in the interview, 479, 480
- observation, 38, 39
- ratings, 54
- time in free association, 364

Rectangular distribution, 77, 91

Reflex, psychogalvanic, *see* Psychogalvanic

Regression, 490

- in rating, 104

Relay, Bean Test, 313, 314

Release, mental, 468

Reliability, 5, 568

- differences in rating traits, 104-108

- loss of, 79

- of Ascendancy-Submission Reaction in Study, 207

- of attitude questionnaires, 231-233

- of Burdick Apperception Test, 552
- of check list, 73

- of conduct knowledge and judgment tests, 284, 285

- of diagnosis, 20, 21, 22

- of Downey Will-Temperament Tests, 343, 344, 351

- of free association classification, 372

- of graphic rating scale, 95

- of Guess Who Test, 73, 74

- of interest questionnaires, 252, 253

- of interviewing, 475-479

- of introversion-extroversion questionnaires, 201

- of judgment tests, 284, 285

- of man-to-man rating, 95

- of measures of deception, 316, 317

- of objective tests, 168

- of observation, 31, 33-36

- of persistence tests, 325

- of personal attitudes tests, 76

Index

595

- Reliability (*continued*)
 - of Pressey X-O Tests, 191, 192
 - of Psychoneurotic Inventory, 185, 186
 - of questionnaire, 161-169
 - of ratings, *see* Rating
 - of self-estimate, 99
 - of Sims Scale of Socio-Economic Status, 552
 - of Studiousness Questionnaire, 227
- Religious
 - and Economic Issues, Watson Survey of Public Opinion of, 216-222, 234, 235
 - education
 - Drew Tests in, 268
 - Indiana Survey of, 302
 - tests, 222, 223
 - Ideas, Chassell Test of, 267
- REMMERS, H. H., 100, 110
- Repetition of stimulus word, 378, 394
- Report, 391, 475
 - personality, 70, 71
- Repression, 488-490
- Reproduction, 366, 377, 379, 391, 394
- Reputation, 562-564
- Research
 - bibliographical, 126
 - interview, 450, 451, 458
 - ratings provide data for, 43, 44
- Resistance, 489, 490, 501
 - body, factor in psychogalvanic reflex, 427
 - breaking, 468
 - capillary, factor in blood pressure, 411
 - electrical, 423, 427
 - in interview, 458
 - meeting, 464-468
 - to opposition, 341
 - to suggestion, 562-564
- Response
 - classification of, in free association, 368-371, 388
 - controlled, 299
 - directed and undirected, 299
 - factors conditioning in free association, 371-373
 - fluidity of, 339
 - measuring, 536
 - natural, 299
 - speed of, 339
- Results of conduct, description of, 12, 13
 - measures of, 12
- Retrospection, 144-146
- Returns on questionnaire, 129, 130
- Rheostat, 423
- RICE, S. A., 216, 233
- RICH, G. J., 437-440
- Right, deviation from accepted idea of, score, 74, 75
- RIKLIN, FR., 361
- Ring, Chinese Test, 324
- Rising action in interviewing, 460, 462-473
- Riva Rocci sphygmomanometer, 408-410
- RIVERS, W. H. R., 491
- ROGERS, C., 300
- ROSANOFF, A. J., 361-363, 370-372, 374, 377, 381, 385-388, 390
- ROSANOFF, I. R., 372, 388
- RUCH, G. M., 168, 186, 279, 289, 308, 310, 346, 348, 351
- RUGG, H. O., 95, 96, 107, 111, 114, 123, 126, 128, 131, 135
- Rules
 - for making graphic rating scale, 65, 66
 - Test, Five, 272
- RYML, B., 61
- RUSK, R. R., 368, 375
- Safe Manipulation Test, 327
- Saliva, 437, 439
 - factors in acidity of, 439
- SALSBERY, P., 469, 470
- Sampling, 5, 474
- SAUDEK, R., 525, 526
- Scale, 19
 - Army Rating, 57-61, 65, 111, 114
 - Behavior Rating, 69
 - Bioet-Simon Intelligence, 15, 267, 268, 351
 - divisions, number of in rating, 78-80
 - for judging intelligence, Pearson's, 53, 54
 - graphic rating, *see* Graphic
 - Group, for Investigating the Emotions, 189, 192
 - Heidbreder Inferiority Attitude Self-rating, 509
 - medical interest, 248, 249
 - men, 61
 - number of items in a rating, 83
 - of Attitude toward Movies, 229, 230
 - of measurement, traits as, 50-52
 - of mental imagery, 53
 - of Occupational Status, Barr, 542-545
 - of Occupations, Taussig, 538
 - Personality Rating, 72
 - rating, *see* Rating
 - Sims, of Socio-Economic Status, *see* Sims
 - Whittier, for Grading Home Conditions, 548, 549

- Schedule
 case study, 15, 16, 556, 557
 interview, 454, 461
 Personality, 183-185, 187
 rating, 42
- Schizoid temperament, 180, 510-512
- Schizophrenia, 180
- Schizophrenic, 509-512
- Scholarship, 12
- SCHNEIDEMÜLL, G., 525
- School
 achievement, prediction of, 388
 effort in, measures of, 333-336
 honesty, 562-564
 marks, 79, 478, 562-564, 566
 Textbooks, Church, Score Card for Measuring, 56, 57
- SCHWEGLER, R. A., 203-205
- SCHWESINGER, G. C., 260, 261, 270, 285, 289
- SCHWESINGER, Ethical-Social Vocabulary Test, 270, 285, 286, 288
- Scientific method, 23
- Scientific synthesis, 560
- Score,
 affectivity, 161, 191, 192
 Card, 55, 56, 83
 for Measuring Church School Textbooks, 56, 57
 for Measuring the Home, 547, 548
 Sims, for Socio-Economic Status, *see* Sims
 criticism of average boy, 74, 75
 deviation from accepted idea of the right, 74, 75
 emotionality, 191-193
 fact, 306, 311
 feeling of difference, 74, 75
 idiosyncrasy, 161, 191-193.
 inferiority, 74, 75
 of amount, 161
 percentile, 169
 quantitative, ratings yield, 45
 self-criticism, 74, 75
 social insight, 74, 75
 superiority and inferiority, 74, 75
- Scoring
 key
 determination of, 249, 250
 by consensus of judgment, 157, 158
 for free association, 390
 questionnaires, methods of, 156-161
- SCOTT, SIR WALTER, 525
- SCOTT, W. D., 57, 61
- Scout, Boy organization, 302
- Seating-plan, in classroom observation, 39
- Secretion, glandular, 440
- Security of contact of electrodes, factor in psychogalvanic reflex, 427
- Selection
 in observation, 30
 item, *see* Item
 of qualities to be rated, 81, 82
 of teachers, 472
- Self
 -confidence, 341, 343, 346
 -criticism, scores, 74, 75
 -estimate, reliability of, 99
 -observation, 18
 -ordinary-ideal rating, 74-76
 -rating, 109-111, 141, 234
 -rating Scale, Heidbreder Inferiority Attitude, 509
- Sense
 of humor, 457
 organs, 24
- Sensitive paper, 403
- Sensory stimulation, factor in psychogalvanic reflex, 429
- Sentence
 Completion Test, C. E. L., 311, 312
 method, 393, 394
 tests, 271-277
- Seriousness of an act, 293
- Service, 12
 tests of, 325-327, 562-564
 intercorrelations of, 326
- Sex, 486-489, 491-493, 496
 and handwriting, 526, 529
 differences
 in attitude, 235
 in complex signs, 380
 in conduct standards, 290
 in introversion-extroversion, 204
 in moral standards, 290
 in questionnaires, 171
 in ratings, 101, 102
 in reaction time, 375, 376
 in testimony, 475
 variations
 of blood pressure with, 411
 of pulse with, 405
- Sexual drives, 487, 488
- Shape of hand, 504, 524
- SHARP, F. C., 268, 269
- Sheet,
 checking, in rating, 46
 personal data, 174
 Pupil Data, 150
 rating, 45, 46
- SHELDON, W. H., 508, 509
- SHEEN, E., 94, 98, 99, 106, 109
- SHUMBERG, M. E., 268, 289
- SHUTTLEWORTH, F. K., 226, 561-566

- Sims, B., 428
 Signs
 complex, 378-381
 of conduct, external, 12, 13, 503-535
 Similarities Test, 270, 286, 288
 SIMON, T., 267
 Simple constellation, 370
 Sims, V. M., 539, 548, 550, 552, 563, 564
 Sims Score Card for Socio-Economic Status, 539, 563, 564
 reliability of, 552
 Sincerity, 456
 Single interview, 474
 Situations, 181, 182
 control of, in observation, 36, 37
 controlled and uncontrolled, 299
 natural, 299, 300
 participation in, in observation, 37
 social, observation of number of, 36
 observation of time spent in, 36
 Size
 factor in blood pressure, 41
 factor in breathing, 418
 of electrodes, factor in psychogalvanic reflex, 427
 of home, 548
 of question blank, 130, 131
 variations of pulse with, 405
 -weight illusion, 318, 319
 Skills, 560
 Slang, Vocabulary of, Schwesinger
 Test of, 261, 262
 SLAVENS, G. S., 288
 SLAWSON, J., 99, 103, 104, 107, 186
 Sleep, factor in breathing, 419
 Small correlations, value of, 559
 SMITH, W. W., 375-377, 379, 430
 Smoked paper, 402, 403
 SNEDDEN, D. S., 454
 SNOW, A. J., 99, 103
 SNYDER, A., 290, 292, 293
 Sociability, 509
 Social
 attitudes
 and interest test, 224, 225
 Questionnaire, 157, 158, 166, 215, 216
 Test of, 167
 contact, observation of, 36
 Ethical, Vocabulary Test, Schwesinger, 270, 285, 286, 288
 importance, basis for selecting items in observation, 31
 information, 235
 insight score, 74, 75
 interest, 389, 390
 level of home, 536
 Social (*continued*)
 selection as basis of item selection, 147, 148
 situations
 observation of number of, 36
 observation of time spent in, 36
 worker, 451, 455, 457, 459, 462, 464, 473, 479, 480
 Socialization, 224
 Socio-economic
 level, measures of, 548, 550-552, 563
 level of the home, 536, 546
 Status,
 Sims Score Card for, *see* Sims
 SOMMER, R., 420
 Sound associations, 368, 369
 Spaces Test, 307
 Spacing observations, 33
 Speaking, factor in breathing, 419
 SPEARMAN, C. E., 33, 95, 201, 253, 307, 310, 313, 325
 Spearman-Brown prophecy formula, 33, 95, 201, 253, 307, 310, 313, 325
 Specific trait, 107
 Speed
 of decision, 329-331, 340, 343, 346, 347
 intercorrelations of measures of, 330
 of movement, 339, 343-346
 of response, 339
 tests, 311, 312, 363
 Sphygmograph, 404, 407
 Sphygmomanometer, 408-410
 Spirit, personnel, 42, 43
 Spirometer, 313
 Spurious validity, 148, 149
 Square, magic number and word, 324
 Test, Magic, 314
 Squares, Nested, Test, 306, 563
 Stability
 emotional, 562-564
 mental, prognosis of, 9
 Standard deviation units, turning ranks into, 91, 93
 Standardization, 568
 of the interview, 6, 454, 455
 vs. versatility, 15, 16
 Standards, 169, 286
 ethical, 290, 291
 occupational intelligence, 540, 541
 of conduct, 19, 20
 Standing broad jump, 313
 Stanford Revision of the Binet-Simon Intelligence Scale, 267, 268, 351
 Stand-outishness, 112
 STARCH, D., 429
 STARLING, E. H., 406, 413
 STARR, H. E., 437

- Statement, true-false, 140
- Statistical
inquiry, 6, 18
methods, 6, 7, 20
treatment of paired comparisons, 86-89, 93
treatment of ranking, 90-93
- Stealing, 303
tests of, 314
- Steps in interviewing, 459-474
- Stethoscope, 409
- Stigmata, 504, 514-516
- Stimulation, sensory, factor in psychogalvanic reflex, 429
- Stimulus word, repetition of, 378, 394
- Stop-watch, 365
- Story, 391, 392
Inhibition Test, 327, 328, 563
Test, 277, 278
- STRANG, R., 263
- STRONG, E. K., JR., 148, 160, 242, 248-250, 253, 254
- Student Information Blank, 226
- Studiosness
Index, 334-336
measures of, 333-336
Questionnaire, 142, 166, 226, 227
reliability of, 227
- Study,
case, *see* Case
individuals, questionnaire to, 125
of the individual, comprehensive, 555-569
- Stupor, catatonic, 431
- STURTEVANT, S. M., 461
- Subjective
associations, 370
definition in rating, 85
- Sublimation, 491
- Submission, ascendancy, *see* Ascendancy
- Substitution Test, Digit-Symbol, 312
- Subtle question, 142, 143
- Sugar, 400, 401
- Suggestibility, tests of, 318-321
- Suggestion
in the interview, 452
resistance to, 562-564
- Suggestive question, 137, 471
- Summary, diagnostic, in the interview, 480
- Superiority score, 74, 75
- Supervision, parental, 548
- Supervisor, 42
- Suppression, 489
- Survey
Watson's, of Public Opinion on Some Religious and Economic Issues, 142, 166, 216-222, 234, 235
- SUTHERLAND, J. W., 78
- Swallowing, variations of pulse with, 406
- Sweat glands, 400, 421, 425-427
- Sweat theory of psychogalvanic phenomena, 425, 426
- SWEET, L., 74, 76
- Symbolism in dreams, 496
- SYMONDS, P. M., 30, 37-39, 73, 76, 79, 98, 112, 113, 142, 154, 155, 157, 166, 171, 182, 185, 215, 226, 232, 235, 251, 255, 334
- Sympathetic nervous system, 400
- Synthesis, scientific, 560
- Systematic
observation, 49
ratings, 114
- Systolic pressure, 409, 410, 432-434
- Syz, H. C., 431
- Tabulating machine, Hollerith, 131, 132
- Tabulation,
forms, 131
in questionnaire, 135, 136
- Tachistoscope, 392
- Tambour, 403, 418
- TANAKA, S., 290, 294
- TANNENBAUM, S. A., 499
- TARCHANOFF, J., 420, 421
- Tarchanoff's phenomena, 421
- Tatbestand, 391-393
- TAUSSIG, F. W., 537-539
- Thurstone Scale of Occupations, 538
- Teacher
marks, 562, 564, 566
selection, 472
- Technique,
double testing, 308-312
duplicating, 303-306
improbable achievement, 306-308
of diagrams, psychoanalysis as, 485
of psychoanalysis, 493-497
psychoanalytic, evaluation of, 498-501
- Techniques
diagnostic, 6, 7, 15
for lessening tension and meeting resistance, 464-468
- Temperament, 507-512, 514, 525, 529, 561
Downey Will-Test, *see* Downey
- melancholic, 393
- Temperature
factor in breathing, 419
of electrodes, factor in psychogalvanic reflex, 428
variations of pulse with, 405
- Tension, emotional, 394
lessening in an interview, 464-468
- TERMAN, L. M., 111, 179, 243, 302, 310, 545

- Terman Group Test of Mental Ability, 351
- Test, 13, 19, 22, 122
 Addition, 311, 312
 American Council Intelligence, 187, 202, 508
 Applications, 275-277, 285-288, 290
 Arguments, 219, 221
 Arithmetic Problems, 311, 312
 Army Alpha Intelligence, 539-541
 Bean Relay, 313, 314
 Bible Knowledge, Advanced, 266
 Binet Intelligence, 15, 267, 268, 351
 Binet Vocabulary, 329
 Books Read, 310
 Burdick Apperception, *see* Burdick
 Cancellation, 311, 312
 Cardboard, 305, 306
 Cause and Effect, 271, 277, 285-288, 290
 C A V D Intelligence, 324
 C A V I Intelligence, 564
 Chassell, of Religious Ideas, 267
 Chinese Ring, 324
 Circles, 307
 Coin-Counting, 314
 Colgate Mental Hygiene, 164, 185, 187, 203
 Completion, *see* Completion
 Comprehensions, 272-276, 285-288, 290
 Coordination, 563, 566
 Cross, 324, 563
 Cross-out, 189-195, 217, 270, 281
 Degree of Truth, 217
 Digit-Symbol Substitution, 312
 Disguised, 14
 Dots, 312
 Duties, 271, 276, 285-288, 290
 Envelopes, 326, 562, 563
 Ethical Discrimination, 260, 267
 Ethical Perception, 268-279
 Fair-mindedness, 161, 216, 217
 Fifteen Puzzle, 307
 Five Rules, 272
 Foresights, 273, 274, 277
 Gates-Strang, of Health Knowledge, 263
 Generalizations, 217, 218
 Guess Who, *see* Guess Who
 Home, 313
 Inference, 220-222
 Information, 311, 312, 562, 563
 Intelligence, *see* Intelligence
 Judgment, *see* Judgment
 Kinetic Will, 322, 323
 Kits, 326, 562, 563, 566
- Test (continued)
 Laycock Biblical Information, 266
 Magic Square, 314
 Manipulation, 327
 Maze, 306, 307
 Mental, *see* Intelligence
 Money Voting, 326, 563
 Moral Judgment, *see* Moral
 Morgan Group, Mental, 351
 Mystery Man, 314
 National Intelligence, 287
 Nested Squares, 306, 563
 new-type, 222
 Number Checking, 311, 312
 objective, *see* Objective
 of Ability to Choose Principles, 277
 of Ability to Weigh Foreseen Consequences, 277, 284, 287, 289
 of delinquency, 390
 of Eye-Control, 331, 332
 of guilt, 382-385
 of lying, 382-385
 of Slang, 261, 262
 Opinion, *see* Opinion
 Opposites, 270, 285, 286, 288
 Orr Good Manners, 264, 270, 286, 552
 Overstatement, 308-310
 Paraffin Paper, 303, 304
 Personal Attitudes, *see* Personal
 Picture Inhibition, 563, 566
 Planted Dime, 314
 Pressey X-O, *see* Pressey
 Principles, 274, 275, 277, 285-288, 290
 Provocations, 272, 273, 276, 285-288, 290
 Puzzle Peg, 307
 Schwesinger, Ethical-Social Vocabulary, 270, 285, 286, 288
 Sentence Completion, C.E.I., 311, 312
 Similarities, 270, 286, 288
 Social Attitudes, 167
 Social Attitudes and Interest, 224, 225
 Spaces, 307
 Story, 277, 278
 Story Inhibition, 327, 328, 563
 Terman Group, of Mental Ability, 351
 Thorndike Intelligence for High School Graduates, 187, 255, 344, 328, 508
 Thorndike-McCall Reading, 235
 time, 272
 Trabue Completion, 329
 Voelker Completion, 303, 304
 Weight Discrimination, 307, 308

Test (*continued*)

Will-Temperament, *see* Downey
Word Consequences, 270, 277, 286,
288

X-O, *see* Pressey

Testimony, 14, 17, 18, 471, 475

accuracy of, 475, 498
age, differences in, 475
error of interpretation in, 475-477
sex differences in, 475

Testing

ability, 298
double, technique, 308-312
methods, 6, 139-142

Tests, 556

achievement, 11
aptitude, 567
athletic performance, 313, 562, 563
Biblical knowledge, 264-267
C. E. I., 277, 284, 286
character and personality, 266
cooperation, 325-327, 563
difficulty, 311, 312
Directions, Woodworth-Wells, 320
Drew, in Religious Education, 268
fear distraction, 332
Health Education, 270, 277
helpfulness, 302
honesty, 150, 302-318, 562-564
I. E. R., 563
Inhibition, *see* Inhibition
intelligence, *see* Intelligence
judgment, *see* Judgment
knowledge, *see* Knowledge
Minnesota College Ability, 508
moral judgment and knowledge,
see Moral
objective, *see* Objective
of cheating, 303-314
of comprehension of proverbs, fa-
bles, and parables, 267, 268
of copying, 303
of introversion, performance, 200
of lying, 315
of opinion, 167, 562-567
of stealing, 314
of Word Knowledge, Thorndike,
261
paper and pencil, 12, 402
performance, 12, 113, 298-360, 565
567, 569
persistence, *see* Persistence
personality and character, 266
Pressey X-O, *see* Pressey
reasoning, 287
recognition, 274, 277, 285-288, 290
religious education, 22, 223
sentence, 271-277
service, *see* Service
speed, 311, 312, 563

Tests (*continued*)

suggestibility, 318-321
trustworthiness, 302
vocabulary, 260-263, 287, 311, 312
word, 270

association, 332

Textbooks, Church School, Score Card
for Measuring, 56, 57

Theory

muscular, of psychogalvanic phe-
nomena, 425, 426
psychoanalytic, 485-493
swcat, of psychogalvanic phenom-
ena, 425, 426
vasomotor, of psychogalvanic phe-
nomena, 425, 426

Therapy, 497

psychoanalytic, 485

Thinking, 223, 224

THOMAS, D. S., 24, 26, 29, 31, 32, 34,
35

THORNDIKE, E. L., 86, 87, 93, 111, 144,
235, 239-243, 255, 261, 287, 302,
324, 328, 364, 508

Thorndike Intelligence Test for High
School Graduates, 187, 255, 324,
328, 508

Thorndike-McCall Reading Test, 235

Thorndike Tests of Word Knowledge,
261

Thorndike Word List, 261, 364

THURSTONE, L. L., 87, 89, 91, 170,
171, 183, 185, 187, 208, 229,
230, 236

THURSTONE, T. G., 183, 185, 187, 208
Thurstone Neurotic Inventory, 208

Thyroid, factor in blood pressure, 413

Time

after eating, factor in acidity of
urine, 439
interval in observation, 33
reaction, *see* Reaction
recording, in free association, 364
spent in social situation, observa-
tion of, 36
test, 272

Titration, formal, method, 438

TOLMAN, E. C., 375

Tone, affective, 377

Tools, 553

TOOPS, H. A., 129, 130

Trabuc Completion Test, 329

Training

in attitudes, 235
of interviewer, 457
raters, 47-50

Traits, 21, 22

as a scale of measurement, 50-52
character, 518, 524
desirable, 101, 109, 110

Traits (*continued*)

- differences in reliability of, in rating, 104-108
- general, 107
- objective, 106, 107
- rating, 52
- ratings on character, 49, 521, 527, 528
- specific, 107
- undesirable, 109, 110
- Transfer, 300, 301
- Transference, 490, 497
- Transformer, 423
- Traube-Hering waves, factor in blood pressure, 412
- Treatment interview, 451
- Trow, W. C., 109, 329-331
- True-false statement, 140
- Trustworthiness tests, 302
- Truth, Degree of, Test, 217
- Truthfulness
 - index, 315
 - of answers, 124
- Tube, Brown, 424
- Types, 367
 - blond(e)-brunette, 504, 522, 523
 - convex and concave, of face, 519, 520
 - criminal, 514-516
 - Kretschmer's, 504, 509-514
 - of distributions yielded by questionnaires, 170
 - of evidence, 12
 - of insanity, 10
 - of interviews, 450, 451
- UNNBROCK, R. S., 339, 343, 344-346, 351
- Uncontrolled situations, 299
- Underrating, 109, 110
- Undesirable traits, 109, 110
- Undirected response, 299
- Uneasiness, 486-488
- Unit of measurement in observation, 33
- Unpleasant
 - affective tone, 377
 - associations, 375
- Urine, 437, 438
 - acidity of, factors in, 439
 - factors in creatinine concentration of, 439, 440
- Use of physiological measures of emotions, 430-437
- U-tube of mercury, 409

Validating items by rating, 147

Validity

- of ascendance-submission questionnaire, 207

Validity (*continued*)

- of attitude questionnaire, 233-237
- of averages, 152, 153
- of interest questionnaire, 253-256
- of introversion-extroversion questionnaires, 202-205
- of judgment tests, 285-294
- of knowledge and judgment tests, 285-294
- of psychoneurotic inventory, 186-188
- of questionnaire, 147-156
- of rating, 108-113
- spurious, 148, 149
- Value of small correlations, 559
- VAN DENBURG, J. K., 548
- Variability
 - in rating, 93
 - of judgments, 102
- Variations of pulse, 405-407
- Vaso
 - constrictor fibers, 414
 - dilator fibers, 414
- Vasomotor theory of psychoagallanic phenomena, 425, 426
- Veracity, 149-156
- VERAGUTH, O., 420, 429, 430
- Verbal
 - organization, 566
 - vs. non-verbal diagnosis, 17, 18
- Verbs, 374
- Verhor, 391
- Versatility vs. standardization, 15, 16
- Versuchsgeschichte, 391, 392
- Vigouzaux, R., 420
- VIOLA, G., 507
- Vocabulary, 471
- Test
 - Binet, 329
 - Ethical-Social, Schwesinger, 270, 285, 286, 288
 - of Slang, 261, 262
 - tests of, 260-263, 287, 311, 312
- Vocational
 - competence, diagnosis of, 8, 10, 11
 - counselors, 11
 - guidance, 516
 - Interest Blank, 169, 250, 253
- VOELKER, P. F., 302-304, 306, 308
- Voelker Completion Test, 303, 304
- Volitional perseveration, 342, 343, 346
- Volume
 - blood, *see* Blood
- Voluntary diagnosis, 14, 15, 568
- interviewing, 457-458
- Voting, Money, Test, 326, 563

WAITE, H., 94

WALLER, A. D., 426

War, World, *see* World

- WASHBURN, M. T., 383
 WASHBURN, J. N., 207, 208
 Water intake, factor in acidity of urine, 439
 WATSON, G. B., 74, 139, 140, 142, 160, 161, 166, 181, 182, 216, 219, 222, 223, 231, 232, 234-236, 266, 268, 269, 284
 Watson Survey of Public Opinion on Some Religious and Economic Issues, 142, 166, 216-222, 234, 235
 WEBB, E., 46, 94, 100, 161
 WEBER, C. O., 193, 289
 WECHSLER, D., 423, 424, 427, 428
 WEHRLIN, K., 388
 Weight, 506, 507
 Discrimination Test, 307, 308
 factor in acidity of urine, 439
 Weighting
 items, 159-161, 248, 249
 McCall-Long method of, 160
 qualities to be rated, 82, 83
 the probable error, 148, 149
 WELLMAN, B., 204
 WELLS, F. L., 51, 100, 102, 103, 111, 320, 365, 367, 371, 373, 374, 377, 378, 388
 WELLS, H. M., 426 428
 WERTHEIMER, F. L., 511, 512
 WERTHEIMER, M., 391, 392, 393
 Wheatstone Bridge, 422, 423
 WHIFFLE, G. M., 318, 475
 WHITLEY, P. C., 164, 165, 191
 WHITLEY, M. T., 265, 266
 Whittier Scale for Grading Home Conditions, 548, 549
 WICKMAN, E. K., 68
 WILEY, L. E., 434, 435
 Will,
 factor in breathing, 419
 factor in psychogalvanic reflex, 429
 Kinetic Test, 322, 323
 Temperament Test, *see* Downey
 WILLETT, G. W., 239, 240
 WILLIAMS, J. H., 548
 WIRKS, E., 350
 Wish, 207, 208
 WOHLGEMUTH, A., 144, 145
 WOOD, B. D., 88, 91, 551
 Woodrow, H., 308, 310, 372
 WOODWORTH, R. S., 146, 154, 164, 168, 169, 171, 174, 178-181, 183, 185-189, 193, 196, 320, 373, 401, 508
 Woodworth Psychoneurotic Inventory, 164, 174-181, 185-189, 195, 196, 208, 401, 508
 Woodworth-Cady Questionnaire, 164
 Woodworth-House Mental Hygiene Inventory, 164, 180
 Woodworth-Mathews Questionnaire, 164, 193
 Woodworth-Wells Directions Tests, 320
 Word
 association tests, 332
 Consequences Test, 270, 277, 286, 288
 Knowledge, Thorndike Tests of, 261
 List, Thorndike, 261, 364
 magic squares, 324
 stimulus, repetition of, 378, 394
 tests, 270
 Words
 abstract, 374
 concrete, 374
 Worker
 case, 555, 560
 clinical, 6, 7
 social, 451, 455, 457, 459, 462, 464, 473, 479, 480
 World War, 114, 174, 187, 337, 433, 488, 489, 539, 565, 566
 WRESCHNER, A., 388
 WUNDT, W., 394
 WYLIE, A. T., 151-153
 WYMAN, J. B., 389

 X-O Tests, Presscy, *see* Presscy

 Yale Psycho-clinic, 37
 Yes-no question, 140
 YOAKUM, C. S., 104, 110, 245

 ZACHRY, C. B., 499
 ZELENY, L. D., 167, 232
 ZIEHEN, Th., 375
 Zones, erogeneous, 487

